

## NCER Working Paper Series

### Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker-Planck Equation

A. Hurn, J. Jeisman and K. Lindsay

Working Paper #9

February 2007

# Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker-Planck Equation

**A.S. Hurn and J. Jeisman**

School of Economics and Finance, Queensland University of Technology

**K.A. Lindsay**

Department of Mathematics, University of Glasgow

## **Abstract**

Many stochastic differential equations (SDEs) do not have readily available closed-form expressions for their transitional probability density functions (PDFs). As a result, a large number of competing estimation approaches have been proposed in order to obtain maximum-likelihood estimates of their parameters. Arguably the most straightforward of these is one in which the required estimates of the transitional PDF are obtained by numerical solution of the Fokker-Planck (or forward-Kolmogorov) partial differential equation. Despite the fact that this method produces accurate estimates and is completely generic, it has not proved popular in the applied literature. Perhaps this is attributable to the fact that this approach requires repeated solution of a parabolic partial differential equation to obtain the transitional PDF and is therefore computationally quite expensive. In this paper, three avenues for improving the reliability and speed of this estimation method are introduced and explored in the context of estimating the parameters of the popular Cox-Ingersoll-Ross and Ornstein-Uhlenbeck models. The recommended algorithm that emerges from this investigation is seen to offer substantial gains in reliability and computational time.

## **Keywords**

stochastic differential equations, maximum likelihood, finite difference, finite element, cumulative distribution function, interpolation.

**JEL Classification** C22, C52

## **Corresponding author**

Joseph Jeisman

School of Economics and Finance

Queensland University of Technology

Brisbane, 4001, Australia

email [j.jeisman@qut.edu.au](mailto:j.jeisman@qut.edu.au)

# 1 Introduction

Stochastic differential equations (SDEs) are commonly used to model the behaviour of important economic variables such as the instantaneous short-term interest rate, asset prices, asset returns and their volatility (see Sundaresan, 2000). The explanatory and/or predictive power of these models depends crucially on the particularisation of the model SDE(s) to real data through the choice of values for their parameters. This choice is complicated by the fact that a closed-form expression for the transitional probability density function (PDF) is rarely available and hence exact maximum-likelihood (EML) estimation of the model parameters is not possible. Consequently, a large number of alternative estimation approaches have been developed. The first major group of these methods seek to retain the ML framework by approximating the transitional PDF by, for example, numerical solutions of the Fokker-Planck (or forward-Kolmogorov) equation (Hurn and Lindsay, 1999; Jensen and Poulsen, 2002), discrete approximations (Florens-Zmirou, 1989; Elerian, 1998; Shoji and Ozaki, 1998; Durham and Gallant, 2002), Hermite polynomial expansions (Aït-Sahalia, 2002; Bakshi and Ju, 2005), or simulation procedures (Pedersen, 1995; Brandt and Santa-Clara, 2002; Hurn, Lindsay and Martin, 2003). Another major group of methods aim to match characteristics of the sample with characteristics of the model by choice of the parameters. Estimators in this group include the general method of moments (GMM) (Chan *et al*, 1992; Hansen and Scheinkman, 1995), indirect estimators (Gourieroux, Montford and Renault, 1993; Gallant and Tauchen, 1996), spectral GMM (based on the characteristic function) (Singleton, 2001; Jiang and Knight, 2002; Chacko and Viciara, 2003), and non-parametric density matching approaches (Aït-Sahalia, 1996). Other approaches, which are harder to categorise, include Bayesian Markov-chain Monte Carlo procedures (Jones, 1998; Elerian, Chib and Shephard, 2001; Eraker, 2001; Roberts and Stramer, 2001) and methods based on estimating-function approximations of the score function (Bibby and Sørensen, 1995; Kessler and Sørensen, 1999; Sørensen, 2000).

Of these competing approaches, maximum-likelihood estimation based on the numerical solution of the Fokker-Planck equation is the closest in spirit to EML. In comparative studies by Jensen and Poulsen (2002) and Hurn *et al* (2006) this approach seems to perform very well; it provides accurate parameter estimates and is completely generic in the sense that, once coded, different models can be handled by trivial modifications of the code. The method does, however, require repeated numerical solution of a partial differential equation and is thus computationally intensive, a fact which may explain why it has received relatively little attention in the financial econometrics literature.

To be specific, the computational effort accumulates through two distinct but interrelated features of the estimation problem. *First*, for each transition in the data set, the transitional PDF must be obtained by numerical solution of the Fokker-Planck parabolic partial differential equation subject to a delta-function initial condition and appropriate boundary conditions. This in itself can be a painstaking undertaking. *Second*, as in all ML estimation algorithms, the transitional PDF must be solved for each transition in the data set and the information from each transition then accumulated

to form an objective function which can be optimised with respect to the parameters. This feature implies that for a data set containing  $N$  transitions, the computational effort required to treat a single transition must be repeated a total of  $N$  times and the process may quickly become impractical as  $N$  grows.

In this paper, three avenues for reducing the computational burden of this approach are introduced and assessed: the first two deal with improving the ease with which the numerical solution of the Fokker-Planck equation may be obtained; and the third is a novel way to reduce the number of transitions for which the numerical procedure needs to be implemented. In this sense the third idea extends beyond the confines of the estimation method considered in this paper as it can be used to accelerate any ML estimation procedure that approximates the transitional PDF. These three suggested improvements are outlined briefly below.

1. The numerical solution of the Fokker-Planck equation requires the accurate resolution of a delta function initial condition for the transitional PDF. As this requires a relatively fine spatial discretisation of the state space, the numerical process requires substantial computational effort. To overcome this problem, a reformulation of the Fokker-Planck equation in terms of the transitional cumulative distribution function (CDF) is developed. The initial condition for the reformulated problem is a step function, which is much easier to resolve numerically than a delta function. This suggests that procedures based on the reformulated problem should be more robust than those based on the traditional formulation, particularly when coarse discretisations of state are used, which will have positive implications for the required computational time.
2. Currently there exists no definitive view of the most appropriate numerical method for solving the Fokker-Planck equation. Hurn and Lindsay (1999) describe a spectral method while Jensen and Poulsen, (2002) favour the method of finite differences. It is hypothesized here that the method of finite elements is better suited to this particular problem because it should allow the spatial nodes to be placed efficiently, thereby allowing comparable accuracy to be obtained at lower computational cost.
3. The final proposal involves dividing the state space into a fixed grid of initial states and obtaining an accurate description of the transitional PDF for each of these initial states by solution of the Fokker-Planck equation. The likelihood associated with every transition in the data set can be computed to good (and uniform) accuracy using Chebyshev interpolation of the known transitional PDFs for the grid of initial states. This method takes advantage of the fact that the SDEs commonly used in economics and finance are time homogeneous, and that the transitional PDF is a continuous function of the initial state. In this way the numerical solution of the Fokker-Planck equation need only be obtained a fixed number of times and need not be repeated for each and every transition in the data set, a substantial saving in the context of the usual sample size in financial applications.

The remainder of this paper is structured as follows. Section 2 provides a brief description of ML estimation of the parameters of SDEs and introduces both the familiar Fokker-Planck equation formulated in terms of the transitional PDF, and the reformulated equation in terms of the transitional CDF. Section 3 describes two finite-difference procedures, one based on the PDF specification of the problem and the other on the CDF specification. Section 4 likewise outlines two finite-element procedures, one for each specification of the problem. Section 5 describes a simulation experiment designed to evaluate the efficacy of the four procedures detailed in Sections 3 and 4. The experiment considers the calculation of log-likelihood for the Cox-Ingersoll-Ross (CIR) and Ornstein-Uhlenbeck (OU) processes<sup>1</sup>. Section 6 outlines the general idea of interpolating transitional PDFs and provides a formal treatment of the proposed interpolation method. It also reports the results of simulation experiments designed to test the accuracy of the proposed interpolation procedure. In Section 7 all of the contributions made in this paper are brought together in a parameter estimation simulation exercise. An empirical application using Treasury Bill interest rate data is presented in Section 8 and Section 9 contains the concluding remarks.

## 2 Maximum-likelihood estimation

A formal statement of the parameter estimation problem to be addressed is as follows. Given the general one-dimensional time-homogeneous SDE

$$dX = \mu(X; \boldsymbol{\theta}) dt + \sqrt{g(X; \boldsymbol{\theta})} dW, \quad (1)$$

the task is to estimate the parameters  $\boldsymbol{\theta}$  of this SDE from a sample of  $(N+1)$  observations  $X_0, \dots, X_N$  of the stochastic process at known times  $t_0, \dots, t_N$ . In the statement of equation (1),  $dW$  is the differential of the Wiener process and the instantaneous drift  $\mu(x; \boldsymbol{\theta})$  and instantaneous diffusion  $g(x; \boldsymbol{\theta})$  are prescribed functions of state.

The ML estimate of  $\boldsymbol{\theta}$  is generated by minimising the negative log-likelihood function of the observed sample, namely

$$-\log \mathcal{L}(\boldsymbol{\theta}) = -\log f_0(X_0 | \boldsymbol{\theta}) - \sum_{k=0}^{N-1} \log f(X_{k+1} | X_k; \boldsymbol{\theta}), \quad (2)$$

with respect to the parameters  $\boldsymbol{\theta}$ . In this expression,  $f_0(X_0 | \boldsymbol{\theta})$  is the density of the initial state and  $f(X_{k+1} | X_k; \boldsymbol{\theta}) \equiv f((X_{k+1}, t_{k+1}) | (X_k, t_k); \boldsymbol{\theta})$  is the value of the transitional PDF at  $(X_{k+1}, t_{k+1})$  for a process starting at  $(X_k, t_k)$  and evolving to  $(X_{k+1}, t_{k+1})$  in accordance with equation (1). Note that the Markovian property of equation (1) ensures that the transitional density of  $X_{k+1}$  at time  $t_{k+1}$  depends on  $X_k$  alone.

---

<sup>1</sup>The CIR (or square-root) process was originally proposed by Feller (1951) but in finance it is more often associated with Cox, Ingersoll and Ross (1985). Similarly, in financial circles the OU process is often associated with Vasicek (1977).

ML estimation relies on the fact that the transitional PDF,  $f(x, t)$ , is the solution of the Fokker-Planck equation

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x} \left( \frac{1}{2} \frac{\partial(g(x; \boldsymbol{\theta})f)}{\partial x} - \mu(x; \boldsymbol{\theta})f \right) \quad (3)$$

satisfying a suitable initial condition and boundary conditions. Suppose, furthermore, that the state space of the problem is  $[a, b]$  and the process starts at  $x = X_k$  at time  $t_k$ . In the absence of measurement error, the initial condition is

$$f(x, t_k) = \delta(x - X_k) \quad (4)$$

where  $\delta$  is the Dirac delta function, and the boundary conditions required to conserve unit density within this interval are

$$\begin{aligned} \lim_{x \rightarrow a^+} \left( \frac{1}{2} \frac{\partial(gf)}{\partial x} - \mu f \right) &= 0, \\ \lim_{x \rightarrow b^-} \left( \frac{1}{2} \frac{\partial(gf)}{\partial x} - \mu f \right) &= 0. \end{aligned} \quad (5)$$

This paper is also concerned with an equivalent statement of this problem in terms of the transitional CDF,  $F(x, t)$ , which is defined in terms of the transitional PDF,  $f(x, t)$ , by

$$F(x, t) = \int_a^x f(u, t) du. \quad (6)$$

When expressed in terms of  $F(x, t)$ , equation (3) takes the form

$$\frac{\partial^2 F}{\partial x \partial t} = \frac{\partial}{\partial x} \left[ \frac{1}{2} \frac{\partial}{\partial x} \left( g \frac{\partial F}{\partial x} \right) - \mu \frac{\partial F}{\partial x} \right] \quad (7)$$

which can be integrated with respect to  $x$  to give

$$\frac{\partial F}{\partial t} = \left[ \frac{1}{2} \frac{\partial}{\partial x} \left( g \frac{\partial F}{\partial x} \right) - \mu \frac{\partial F}{\partial x} \right] + C(t). \quad (8)$$

where  $C(t)$  is an arbitrary function of integration. The boundary conditions for this equation require that  $F(a, t) = 0$  and  $F(b, t) = 1$  which in turn require that  $C(t) = 0$ . Therefore  $F(x, t)$  satisfies the partial differential equation

$$\frac{\partial F}{\partial t} = \frac{1}{2} \frac{\partial}{\partial x} \left( g \frac{\partial F}{\partial x} \right) - \mu \frac{\partial F}{\partial x} \quad (9)$$

with Dirichlet boundary conditions  $F(a, t) = 0$  and  $F(b, t) = 1$ . The initial condition  $F(x, t_k)$  for a transition from  $(X_k, t_k)$  is constructed from the definition (6) to obtain

$$F(x, t_k) = \begin{cases} 0 & x < X_k, \\ 1/2 & x = X_k, \\ 1 & x > X_k. \end{cases} \quad (10)$$

One important advantage of this approach is that the delta function initial condition required in the computation of transitional PDF is now replaced by a step function initial condition in the computation of the transitional CDF. The latter has a precise numerical representation whereas the delta function (4) must be approximated.

Procedures for solving equations (3) and (9) numerically using finite-difference and finite-element methods are described in detail in the following two sections.

### 3 Finite-difference procedure

The finite-difference procedure is based on a discretisation of state space into  $n$  uniform sub-intervals of length  $\Delta_x = (b - a)/n$  and a discretisation of the time interval  $[t_k, t_{k+1}]$  into  $m$  uniform sub-intervals of duration  $\Delta_t = (t_{k+1} - t_k)/m$ . Let the nodes of the finite-difference scheme be denoted by  $x_p = a + p \Delta_x$  where  $p$  is an integer satisfying  $0 \leq p \leq n$ , let  $t_k = t_{k,0}, t_{k,1}, \dots, t_{k,m} = t_{k+1}$  where  $t_{k,q} = t_k + q \Delta_t$  denote the subdivision of  $[t_k, t_{k+1}]$  into intervals of duration  $\Delta_t$  and let  $f_p^{(q)} = f(x_p, t_{k,q})$  be the value of the transitional PDF at  $x_p$  at time  $t_{k,q}$ .

#### 3.1 Transitional PDF specification

Integration of equation (3) over  $[t_{k,q}, t_{k,q+1}]$  gives

$$f(x, t_{k,q+1}) - f(x, t_{k,q}) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \left( g(x) \int_{t_{k,q}}^{t_{k,q+1}} f(x, t) dt \right) - \frac{\partial}{\partial x} \left( \mu(x) \int_{t_{k,q}}^{t_{k,q+1}} f(x, t) dt \right). \quad (11)$$

In terms of the auxiliary variables

$$\phi_p = \int_{t_{k,q}}^{t_{k,q+1}} f(x_p, t) dt,$$

equation (11) has finite difference approximation

$$f_p^{(q+1)} - f_p^{(q)} = \frac{g_{p+1} \phi_{p+1} - 2g_p \phi_p + g_{p-1} \phi_{p-1}}{2\Delta_x^2} - \frac{\mu_{p+1} \phi_{p+1} - \mu_{p-1} \phi_{p-1}}{2\Delta_x}$$

which may be regrouped to give

$$f_p^{(q+1)} - f_p^{(q)} = \left( \frac{g_{p-1} + \mu_{p-1} \Delta_x}{2\Delta_x^2} \right) \phi_{p-1} - \frac{g_p}{\Delta_x^2} \phi_p + \left( \frac{g_{p+1} - \mu_{p+1} \Delta_x}{2\Delta_x^2} \right) \phi_{p+1}.$$

The trapezoidal quadrature is now used to approximate  $\phi_p$  by the formula

$$\phi_p = \frac{\Delta_t}{2} (f_p^{(q+1)} + f_p^{(q)}) + O(\Delta_t)^3$$

so that the final finite-difference representation of equation (3) simplifies to give

$$\begin{aligned} & - \left[ g_{p-1} + \mu_{p-1} \Delta_x \right] f_{p-1}^{(q+1)} + \left[ \frac{4}{r} + 2g_p \right] f_p^{(q+1)} - \left[ g_{p+1} - \mu_{p+1} \Delta_x \right] f_{p+1}^{(q+1)} \\ & = \left[ g_{p-1} + \mu_{p-1} \Delta_x \right] f_{p-1}^{(q)} + \left[ \frac{4}{r} - 2g_p \right] f_p^{(q)} + \left[ g_{p+1} - \mu_{p+1} \Delta_x \right] f_{p+1}^{(q)} \end{aligned} \quad (12)$$

where  $r = \Delta_t / \Delta_x^2$  is the Courant number. The procedure used to construct equation (12) is essentially the Crank-Nicolson algorithm, and it is well known that this algorithm exhibits robust numerical properties, for example, it is stable and numerically consistent. Expression (12) forms the core of the finite-difference representation of equation (3). It suggests that the transitional PDF can be

integrated forward through time from a given initial distribution by repeated solution of a tri-diagonal system of equations.

As has already been remarked, the initial condition for the transitional PDF is a delta function and is therefore not representable within the framework of the finite-difference method. Jensen and Poulsen (2002) suggest that this difficulty can be circumvented by starting the finite-difference algorithm with a specification of the distribution of transitional density at  $(t_k + \Delta_t)$  based on the assumption that the transitional density at this time may be approximated by the normal distribution with mean value  $X_k + \mu(X_k; \theta)\Delta_t$  and variance  $g(X_k; \theta)\Delta_t$ . The main drawback of this approximation is that once  $\Delta_t$  is chosen and the initial state is known, the diffusion occurring over the time interval  $\Delta_t$  from the true initial condition determines the size of the interval of state space over which the transitional PDF is significantly different from zero. The resolution  $\Delta_x$  of state space must now be chosen to be sufficiently small so as to guarantee that a reasonable number of nodes (say a dozen) lie within this interval of non-zero transitional PDF. Moreover, once a suitable value of  $\Delta_x$  is chosen, this discretisation interval must be applied to the entire state space. In practice, this requirement means that  $\Delta_x = O(\sqrt{\Delta_t})$ .

A final crucial aspect of the finite-difference algorithm is the incorporation of the boundary conditions into the first and last equations in the system. Recall that the solution is sought in the finite interval  $[x_0, x_n]$ . For many SDEs of type (1), the sample space is the semi-infinite interval  $(0, \infty)$  so that the drift and diffusion specifications will often satisfy  $g(x_0) = 0$  and  $\mu(x_0) > 0$ . Under these conditions the boundary condition at  $x = x_0$  is equivalent to the condition  $f(x_0, t) = 0$ , that is, no density can accumulate at the boundary  $x = x_0$ . However, no equivalent simplification exists at the boundary  $x = x_n$  which must be chosen to be suitably large, but finite<sup>2</sup>. The derivation of the boundary condition at  $x = x_n$  is now described.

The backward-difference representation of the boundary condition (5) at  $x = x_n$  is

$$\frac{1}{2} \left( \frac{3g_n f_n^{(q)} - 4g_{n-1} f_{n-1}^{(q)} + g_{n-2} f_{n-2}^{(q)}}{2\Delta_x} \right) - \mu_n f_n^{(q)} + O(\Delta_x)^2 = 0. \quad (13)$$

These terms are regrouped and the truncation error ignored to obtain

$$(3g_n - 4\mu_n \Delta_x) f_n^{(q)} - 4g_{n-1} f_{n-1}^{(q)} + g_{n-2} f_{n-2}^{(q)} = 0. \quad (14)$$

This boundary condition is now used at  $(t_k + q\Delta_t)$  and  $(t_k + (q+1)\Delta_t)$  to eliminate  $f_n^{(q)}$  and  $f_n^{(q+1)}$  respectively from equation (12) evaluated at  $p = n-1$ . The final result is

$$P f_{n-2}^{(q+1)} - (Q - R) f_{n-1}^{(q+1)} = -P f_{n-2}^{(q)} + (Q + R) f_{n-1}^{(q)} \quad (15)$$

where

$$P = g_{n-2} (3\mu_n \Delta_x - 2g_n) - \mu_{n-2} \Delta_x (3g_n - 4\mu_n \Delta_x),$$

$$Q = g_{n-1} (4\mu_n \Delta_x - 2g_n), \quad R = \frac{4}{r} (3g_n - 4\mu_n \Delta_x).$$

---

<sup>2</sup>In the applications here,  $x_n$  is chosen to be the maximum of the sample plus the range of the sample.



When it is not possible to assume that  $f(x_0, t) = 0$ , the lower boundary condition can be derived using a similar procedure. The result is an identical expression to equation (15) but with the subscripts  $n$ ,  $n - 1$  and  $n - 2$  replaced by 0, 1 and 2 respectively, and the negative sign between the two terms in  $P$  replaced by a positive sign.

The final specification of the finite-difference procedure can be expressed in matrix form as

$$A_L \mathbf{f}^{(q+1)} = A_R \mathbf{f}^{(q)}$$

where  $A_L$  and  $A_R$  are tri-diagonal matrices of dimension  $(n - 1) \times (n - 1)$  and  $\mathbf{f}^{(q)}$  is the  $(n - 1)$  dimensional vector containing the values of the transitional PDF at the (internal) nodes  $x_1, \dots, x_{n-1}$  at time  $(t_k + q\Delta_t)$ . In the semi-infinite case, the first equation in this tri-diagonal system is

$$\left[\frac{4}{r} + 2g_1\right]f_1^{(q+1)} - \left[g_2 - \mu_2 \Delta_x\right]f_2^{(q+1)} = \left[\frac{4}{r} - 2g_1\right]f_1^{(q)} + \left[g_2 - \mu_2 \Delta_x\right]f_2^{(q)}, \quad (16)$$

which is the particularisation of the general equation (12) at  $x_1$  taking account of the requirement that  $f(x_0, t) = 0$ . In all other cases the first equation would be a lower boundary condition of the type described in the discussion following equation (15). In all cases there then follow  $(n - 3)$  equations with general form (12) in which the index  $p$  takes values from  $p = 2$  to  $p = n - 2$ , followed finally by equation (15). Note that the resulting tri-diagonal system is solved for the transitional PDF at the nodes  $x_1, \dots, x_{n-1}$ . In the semi-infinite case the transitional PDF at  $x_0$  is known *a priori* to be zero, while the transitional PDF at  $x_n$  is obtained directly from relation (14). In other cases the transitional PDF at  $x_0$  is obtained from an almost identical relation to equation (14) but with the subscripts  $n$ ,  $n - 1$  and  $n - 2$  replaced by 0, 1 and 2 respectively.

### 3.2 Transitional CDF specification

The finite-difference representation of equation (9) is constructed by noting that this equation can be re-expressed in the form

$$\frac{\partial F}{\partial t} = \frac{1}{4} \left[ \frac{\partial^2(gF)}{\partial x^2} + g \frac{\partial^2 F}{\partial x^2} - F \frac{\partial^2 g}{\partial x^2} \right] - \mu \frac{\partial F}{\partial x}. \quad (17)$$

The motivation for this manipulation stems from the fact that the central-difference formula for a second order derivative is straightforward. The procedure used to derive equation (12) from equation (3) via equation (11) is repeated for equation (17). The calculation is routine and so the details are suppressed. Let  $F_p^{(q)} = F(x_p, t_k + q\Delta_t)$  denote the value of the CDF at  $(x_p, t_k + q\Delta_t)$ , then the finite-difference approximation of equation (17) is

$$\begin{aligned} & - \left[ g_{p-1} + g_p + 2\mu_p \Delta_x \right] F_{p-1}^{(q+1)} + \left[ \frac{8}{r} + (g_{p-1} + 2g_p + g_{p+1}) \right] F_p^{(q+1)} \\ & - \left[ g_p + g_{p+1} - 2\mu_p \Delta_x \right] F_{p+1}^{(q+1)} = \left[ g_{p-1} + g_p + 2\mu_p \Delta_x \right] F_{p-1}^{(q)} \\ & + \left[ \frac{8}{r} - (g_{p-1} + 2g_p + g_{p+1}) \right] F_p^{(q)} + \left[ g_p + g_{p+1} - 2\mu_p \Delta_x \right] F_{p+1}^{(q)}. \end{aligned} \quad (18)$$

However the boundary conditions assert that  $F_0^{(q)} \equiv 0$  and  $F_n^{(q)} \equiv 1$ , and therefore equations (18) can be expressed in matrix form

$$A_L \mathbf{F}^{(q+1)} = A_R \mathbf{F}^{(q)} + \mathbf{B}$$

where  $A_L$  and  $A_R$  are tri-diagonal matrices of dimension  $(n-1) \times (n-1)$ ,  $\mathbf{B}$  is a constant vector of dimension  $(n-1)$  which differs from the zero vector only in its last entry and  $\mathbf{F}^{(q)}$  is the  $(n-1)$  dimensional vector containing the values of the transitional CDF at the (internal) nodes  $x_1, \dots, x_{n-1}$  at time  $t_{k,q}$ .

In the practical implementation of this procedure, there is a natural initial condition given in equation (10), for which there is no equivalent statement in the transitional PDF formulation of the problem. Of course, the Jensen and Poulsen (2002) approximation can also be used in the transitional CDF approach by initialising  $F_1, \dots, F_{n-1}$  using the CDF of the normal distribution. The value of the transitional PDF at  $(X_{k+1}, t_{k+1})$  is estimated by numerical differentiation (using difference formulae) of the transitional CDF at the nodes to the left and right of  $X_{k+1}$  followed by linear interpolation of these values to find the required transitional density.

## 4 Finite-element procedure

The finite-element approach also requires the discretisation of state space and time. As with the finite-difference procedure, the time interval  $[t_k, t_{k+1}]$  is subdivided into  $m$  uniformly-spaced intervals of length  $\Delta_t$  and the state space of the problem is again subdivided into a sequence of  $n$  connected intervals, or elements,  $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$ . However, a finite-element procedure differs from a finite-difference procedure in the respect that the nodes  $x_0, \dots, x_n$  need not be uniformly spaced in state space. By comparison with a finite-difference algorithm, a finite-element scheme to solve the Fokker-Planck equation will position proportionately more nodes in the main body of the transitional PDF and proportionately less nodes in the tails of the transitional PDF. The distribution of nodes in the finite-element algorithm is determined by a user-supplied rule<sup>3</sup> which may itself be problem dependent.

Once the elements are laid out, a family of basis functions, say  $\psi_0(x), \dots, \psi_n(x)$ , is superimposed on the elements such that  $\psi_j(x)$  is a continuous function of  $x$  with support  $(x_{j-1}, x_{j+1})$ . Clearly  $\psi_0$  and  $\psi_n$  are each defined over a single element. In this paper, the basis function  $\psi_j(x)$  is taken to be the simple triangular or ‘‘tent’’ basis function

$$\psi_j(x) = \begin{cases} \frac{x - x_{j-1}}{x_j - x_{j-1}} & x \in [x_{j-1}, x_j] \\ \frac{x_{j+1} - x}{x_{j+1} - x_j} & x \in [x_j, x_{j+1}] \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

---

<sup>3</sup>In the applications here  $X_k$  is set as node  $x_i$  where  $i = \text{ceil}(nX_k/(b-a))$ . Node  $x_j$  is then set as  $x_i - (X_k - a)/\sqrt{j+1}$  for  $j = 1, \dots, i-1$  or  $x_i + (b - X_k)/\sqrt{n-j-1}$  for  $j = i+1, \dots, n$ .

Figure 1 illustrates the general structure of the basis functions.

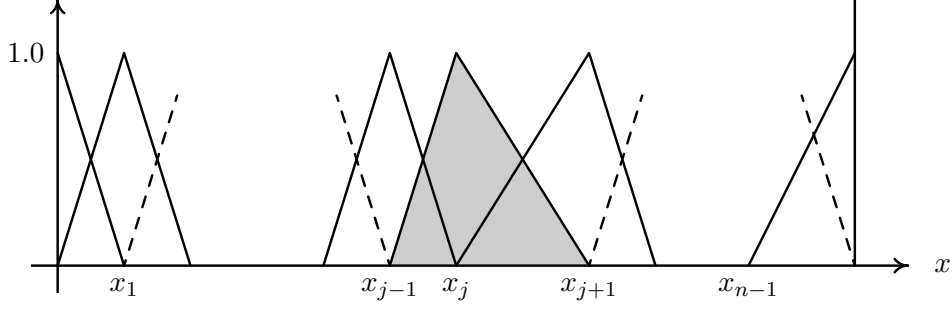


Figure 1: Family of basis functions  $\psi_0(x), \psi_1(x), \dots, \psi_n(x)$  with member  $\psi_j(x)$  shaded.

The central idea of the finite-element procedure is that any arbitrary function of state, say  $h(x)$ , can be represented by a series expansion of the type

$$\hat{h}(x) = \sum_{i=0}^n h_i \psi_i(x) \quad (20)$$

where the coefficients  $h_0, \dots, h_n$  are constants to be determined. Since  $h(x)$  and its finite-element expansion  $\hat{h}(x)$  cannot be expected to agree for all values of  $x$ , the coefficients  $h_0, \dots, h_n$  are often chosen to minimise

$$E(h_0, \dots, h_n) = \int_a^b [h(x) - \hat{h}(x)]^2 dx = \int_a^b \left[ h(x) - \sum_{i=0}^n h_i \psi_i(x) \right]^2 dx. \quad (21)$$

Evidently equation (21) is minimised by requiring  $h_0, \dots, h_n$  to satisfy the  $(n+1)$  conditions

$$\frac{\partial E}{\partial h_j} = -2 \int_a^b \left[ h(x) - \sum_{i=0}^n h_i \psi_i(x) \right] \psi_j(x) dx = 0, \quad j = 0, \dots, n,$$

which may be reorganised into the more familiar form

$$\sum_{i=0}^n h_i \int_a^b \psi_i(x) \psi_j(x) dx = \int_a^b h(x) \psi_j(x) dx, \quad j = 0, \dots, n. \quad (22)$$

#### 4.1 Transitional PDF specification

Suppose now that the transitional PDF has finite-element expansion

$$\hat{f}(x, t) = \sum_{i=0}^n f_i(t; \boldsymbol{\theta}) \psi_i(x). \quad (23)$$

Clearly, the coefficients  $f_i(t)$  can be determined directly by a least squares fit only when the transitional PDF is known. When the transitional PDF is unknown, as is the case here, instead the finite-element representation of the Fokker-Planck operator

$$\mathcal{K}(f) = \frac{\partial f}{\partial t} - \frac{\partial}{\partial x} \left( \frac{1}{2} \frac{\partial(g(x; \boldsymbol{\theta})f)}{\partial x} - \mu(x; \boldsymbol{\theta})f \right) \quad (24)$$

is required to be zero. Equation (22) now indicates that  $f_0(t), \dots, f_n(t)$  must be chosen to satisfy

$$\int_a^b \mathcal{K}(\hat{f}) \psi_j(x) dx = 0, \quad j = 0, \dots, n. \quad (25)$$

Straight forward manipulation of this equation involving one integration by parts and the use of the zero-flux boundary conditions (5) yields

$$\int_a^b \frac{\partial \hat{f}}{\partial t} \psi_j(x) dx = - \int_a^b \left( \frac{1}{2} \frac{\partial(g(x; \boldsymbol{\theta}) \hat{f})}{\partial x} - \mu(x; \boldsymbol{\theta}) \hat{f} \right) \frac{d\psi_j(x)}{dx} dx. \quad (26)$$

To make further progress, the drift and diffusion functions are given the finite-element expansions

$$\hat{\mu}(x, t) = \sum_{s=0}^n \mu_s(t; \boldsymbol{\theta}) \psi_s(x), \quad \hat{g}(x, t) = \sum_{s=0}^n g_s(t; \boldsymbol{\theta}) \psi_s(x) \quad (27)$$

which are now substituted into equation (26) to obtain

$$\begin{aligned} \sum_{i=0}^n \frac{df_i}{dt} \int_a^b \psi_i \psi_j dx &= \sum_{i=0, s=0}^n f_i \mu_s \int_a^b \psi_i \psi_s \frac{d\psi_j}{dx} dx \\ &\quad - \frac{1}{2} \sum_{i=0}^n \sum_{s=0}^n f_i g_s \int_a^b \left( \psi_i \frac{d\psi_s}{dx} \frac{d\psi_j}{dx} + \psi_s \frac{d\psi_i}{dx} \frac{d\psi_j}{dx} \right) dx, \end{aligned} \quad j = 0, 1 \dots n. \quad (28)$$

Each integer from  $j = 0$  to  $j = n$  inclusive contributes one ordinary differential equation to give  $(n + 1)$  equations in total. The integrals arising in equation (28) can be evaluated exactly for the basis functions defined by equation (19). The details of these calculations are straightforward but tedious and are given in Appendix A. Essentially the interval  $[a, b]$  of integration is expressed as a union of elements, and the values of the integrals are determined by summing the contributions made by the individual elements. The calculations in Appendix A may now be used to expand the summations in equations (28) to obtain the final equations

$$(x_1 - x_0) \left( 2 \frac{df_0}{dt} + \frac{df_1}{dt} \right) = - \left( 2\mu_0 + \mu_1 + \frac{3g_0}{(x_1 - x_0)} \right) f_0 - \left( \mu_0 + 2\mu_1 - \frac{3g_1}{(x_1 - x_0)} \right) f_1 \quad (29)$$

in the case in which  $j = 0$ , the equations

$$\begin{aligned} (x_j - x_{j-1}) \frac{df_{j-1}}{dt} + 2(x_{j+1} - x_{j-1}) \frac{df_j}{dt} + (x_{j+1} - x_j) \frac{df_{j+1}}{dt} &= \left( 2\mu_{j-1} + \mu_j + \frac{3g_{j-1}}{(x_j - x_{j-1})} \right) f_{j-1} \\ &\quad - \left( \mu_{j+1} - \mu_{j-1} + \frac{3g_j(x_{j+1} - x_{j-1})}{(x_j - x_{j-1})(x_{j+1} - x_j)} \right) f_j - \left( \mu_j + 2\mu_{j+1} - \frac{3g_{j+1}}{(x_{j+1} - x_j)} \right) f_{j+1} \end{aligned} \quad (30)$$

in the case in which  $0 < j < n$ , and finally the equation

$$\begin{aligned} (x_n - x_{n-1}) \left( \frac{df_{n-1}}{dt} + 2 \frac{df_n}{dt} \right) &= \left( 2\mu_{n-1} + \mu_n + \frac{3g_{n-1}}{(x_n - x_{n-1})} \right) f_{n-1} \\ &\quad + \left( \mu_{n-1} + 2\mu_n - \frac{3g_n}{(x_n - x_{n-1})} \right) f_n \end{aligned} \quad (31)$$

in the case in which  $j = n$ . These equations can be expressed in the matrix form

$$A_L \frac{d\mathbf{V}}{dt} = A_R \mathbf{V} \quad (32)$$

where  $A_L$  and  $A_R$  are  $(n+1) \times (n+1)$  tri-diagonal matrices and  $\mathbf{V}$  is the  $(n+1)$  dimensional column vector with  $i$ -th row containing the value of the coefficient  $f_i$ . Let  $\mathbf{V}^{(q)}$  be the solution of equation (32) at time  $t_{k,q} = t_k + q \Delta_t$ , then integration of equations (32) over  $[t_{k,q}, t_{k,q+1}]$  gives

$$A_L(\mathbf{V}^{(q+1)} - \mathbf{V}^{(q)}) = \frac{\Delta_t}{2} [A_R \mathbf{V}^{(q)} + A_R \mathbf{V}^{(q+1)}] + O(\Delta_t)^3 \quad (33)$$

where the fundamental theorem of calculus has been used to integrate the time derivative on the left hand side of equation (32) and the trapezoidal rule has been used to estimate the integral of the right hand side of equation (32). Collecting like terms together yields

$$\left[ A_L - \frac{\Delta_t}{2} A_R \right] \mathbf{V}^{(q+1)} = \left[ A_L + \frac{\Delta_t}{2} A_R \right] \mathbf{V}^{(q)} \quad (34)$$

where terms of order  $(\Delta_t)^3$  have been ignored. In principle, the starting values of the finite-element coefficients are used to initialise  $\mathbf{V}^{(0)}$ . The values of the finite-element coefficients at  $t_{k+1}$ , namely  $\mathbf{V}^{(m)}$ , are then determined by means of  $m$  repeated applications of equation (34) in which the intermediate coefficient values  $\mathbf{V}^{(1)}, \dots, \mathbf{V}^{(m-1)}$  are generated and discarded. Although each step of this iteration is accurate to order  $(\Delta_t)^3$ , the final values of the finite-element coefficients are determined to an accuracy of order  $(\Delta_t)^2$ .

The starting values for the finite-element coefficients can be obtained by recognising that  $\hat{f}(x_i, t) = f_i$  because at node  $x_i$  the basis function  $\psi_i$  is unity while all of the other basis functions are zero. This suggests that the finite-element coefficients can be initialised in the same manner as the transitional PDF. The Gaussian approximation of the transitional PDF at time  $t_k + \Delta_t$  proposed by Jensen and Poulsen (2002) can therefore be used to initialise  $\mathbf{V}^{(1)}$  and the final values of the finite-element coefficients are constructed by  $(m-1)$  applications of equation (34).

## 4.2 Transitional CDF specification

Suppose that the transitional CDF has the finite-element expansion

$$\hat{F}(x, t) = \sum_{i=0}^n F_i(t; \boldsymbol{\theta}) \psi_i(x). \quad (35)$$

Once again, the coefficients  $F_0(t), \dots, F_n(t)$  of this expansion cannot be determined by a least squares fit, but instead are obtained indirectly by fitting expression (35) to equation (9) in the sense that the coefficients  $F_0(t), \dots, F_n(t)$  are required to ensure that

$$\int_a^b \left( \frac{\partial \hat{F}}{\partial t} - \frac{1}{2} \frac{\partial}{\partial x} \left( g \frac{\partial \hat{F}}{\partial x} \right) + \mu \frac{\partial \hat{F}}{\partial x} \right) \psi_j(x) dx = 0, \quad j = 1, \dots, (n-1). \quad (36)$$

Notice that the cases  $j = 0$  and  $j = n$  are excluded from equation (36). This is because  $F_0(t) = 0$  and  $F_n(t) = 1$  in view of the requirement that expression (36) should satisfy the boundary conditions  $\hat{F}(a, t) = 0$  and  $\hat{F}(b, t) = 1$ . Furthermore, integration by parts yields

$$\begin{aligned} \int_a^b \frac{\partial}{\partial x} \left( g \frac{\partial \hat{F}}{\partial x} \right) \psi_j dx &= \left[ g \frac{\partial \hat{F}}{\partial x} \psi_j \right]_a^b - \int_a^b g \frac{\partial \hat{F}}{\partial x} \frac{d\psi_j}{dx} dx \\ &= - \int_a^b g \frac{\partial \hat{F}}{\partial x} \frac{d\psi_j}{dx} dx \end{aligned} \quad j = 1, \dots, (n-1)$$

where the boundary contribution has been eliminated on the basis that  $\psi_j(a) = \psi_j(b) = 0$  for all  $j = 1, \dots, (n-1)$ . In conclusion, the coefficients  $F_1(t), \dots, F_{n-1}(t)$  are chosen to satisfy

$$\int_a^b \left( \frac{\partial \hat{F}}{\partial t} + \mu \frac{\partial \hat{F}}{\partial x} \right) \psi_j(x) dx + \int_a^b \frac{g}{2} \frac{\partial \hat{F}}{\partial x} \frac{d\psi_j}{dx} dx = 0, \quad j = 1, \dots, (n-1). \quad (37)$$

As with the finite-element treatment of the transitional PDF, the drift and diffusion functions are given the finite-element expansions

$$\hat{\mu}(x, t) = \sum_{k=0}^n \mu_k(t; \boldsymbol{\theta}) \psi_k(x), \quad \hat{g}(x, t) = \sum_{k=0}^n g_k(t; \boldsymbol{\theta}) \psi_k(x). \quad (38)$$

Equation (37) now leads to the final conclusion that the coefficients  $F_1(t), \dots, F_{n-1}(t)$  satisfy

$$\begin{aligned} \sum_{i=j-1}^{j+1} \frac{dF_i}{dt} \int_a^b \psi_i \psi_j dx &= - \sum_{i=j-1}^{j+1} \sum_{k=j-1}^{j+1} F_i \mu_k \int_a^b \psi_j \psi_k \frac{d\psi_i}{dx} dx \\ &\quad - \frac{1}{2} \sum_{i=j-1}^{j+1} \sum_{k=j-1}^{j+1} F_i g_k \int_a^b \psi_k \frac{d\psi_i}{dx} \frac{d\psi_j}{dx} dx, \quad j = 1, \dots, (n-1), \end{aligned} \quad (39)$$

where the values of the integrals in these equations may be computed using the results provided in Appendix A. It is a straight forward but tedious matter to verify that equations (39) simplify to give

$$\begin{aligned} 2(x_2 - x_0) \frac{dF_1}{dt} + (x_2 - x_1) \frac{dF_2}{dt} &= \\ - \left( \mu_2 + \mu_0 + \frac{3(g_0 + g_1)}{2(x_1 - x_0)} + \frac{3(g_1 + g_2)}{2(x_2 - x_1)} \right) F_1 &- \left( 2\mu_1 + \mu_2 - \frac{3(g_1 + g_2)}{2(x_2 - x_1)} \right) F_2 \end{aligned} \quad (40)$$

when  $j = 1$  and the equations

$$\begin{aligned} (x_j - x_{j-1}) \frac{dF_{j-1}}{dt} + 2(x_{j+1} - x_{j-1}) \frac{dF_j}{dt} + (x_{j+1} - x_j) \frac{dF_{j+1}}{dt} &= \\ \left( \mu_{j-1} + 2\mu_j + \frac{3(g_{j-1} + g_j)}{2(x_j - x_{j-1})} \right) F_{j-1} &- \left( \mu_{j-1} - \mu_{j+1} + \frac{3(g_{j-1} + g_j)}{2(x_j - x_{j-1})} + \frac{3(g_j + g_{j+1})}{2(x_{j+1} - x_j)} \right) F_j \\ - \left( 2\mu_j + \mu_{j+1} - \frac{3(g_j + g_{j+1})}{2(x_{j+1} - x_j)} \right) F_{j+1} & \end{aligned} \quad (41)$$

when  $1 < j < n-1$ , and finally the equation

$$\begin{aligned} (x_{n-1} - x_{n-2}) \frac{dF_{n-2}}{dt} + 2(x_n - x_{n-2}) \frac{dF_{n-1}}{dt} &= \left( \mu_{n-2} + 2\mu_{n-1} + \frac{3(g_{n-2} + g_{n-1})}{2(x_{n-1} - x_{n-2})} \right) F_{n-2} \\ - \left( \mu_{n-2} - \mu_n + \frac{3(g_{n-2} + g_{n-1})}{2(x_{n-1} - x_{n-2})} + \frac{3(g_{n-1} + g_n)}{2(x_n - x_{n-1})} \right) F_{n-1} &- \left( 2\mu_{n-1} + \mu_n - \frac{3(g_{n-1} + g_n)}{2(x_n - x_{n-1})} \right) F_n \end{aligned} \quad (42)$$

when  $j = (n - 1)$ . The procedure to solve these equations is identical to that used to solve the corresponding equations arising in the finite-element procedure to determine the transitional PDF and so the specific details are omitted. The finite-element coefficients can either be initialised at time  $t_k$  using the step function initial condition (10) or by using the CDF of the Normal distribution at time  $t_k + \Delta_t$ .

**Computation of PDF** The final stage of the transitional CDF approach is to recover the transitional PDF from the finite-element expansion for the transitional CDF. In theory the transitional PDF can be obtained by differentiating the transitional CDF. However, the finite-element representation of the transitional CDF is not continuously differentiable, and therefore a least-squares fitting procedure must be used in order to obtain a continuous expression for the transitional PDF from the transitional CDF. The aim of this procedure is to construct the finite-element representation

$$\hat{f}(x, t) = \sum_{i=0}^n f_i(t) \psi_i(x)$$

of the transitional PDF by choosing the coefficients  $f_0, \dots, f_n$  to minimise

$$\Phi(f_0, \dots, f_n) = \int_a^b \left[ \hat{f}(x, t) - \frac{d\hat{F}}{dx} \right]^2 dx$$

This is achieved by choosing  $f_0, \dots, f_n$  to be the solution of the linear equations

$$\sum_{i=0}^n f_i \int_a^b \psi_i \psi_j dx = \sum_{i=0}^n F_i \int_a^b \frac{d\psi_i}{dx} \psi_j dx, \quad j = 0, \dots, n. \quad (43)$$

The computation of the integrals in this equation is described in Appendix A, but the final result is that the values of the coefficients  $f_0, \dots, f_n$  are obtained as the solution of the simultaneous equations  $T\mathbf{f} = \mathbf{b}$  where  $T$  is an  $(n + 1) \times (n + 1)$  symmetric tri-diagonal matrix,  $\mathbf{f} = (f_0, \dots, f_n)^T$  and  $\mathbf{b}$  is a vector of dimension  $(n + 1)$  which is determined by knowledge of  $F_0, \dots, F_n$ .

## 5 Log-likelihood computation

The efficacy of the numerical procedures depends on the accuracy with which they can compute the log-likelihood of a sample. Accordingly, this section outlines a simulation experiment designed to investigate the accuracy of log-likelihood computation for the finite-difference and finite-element procedures described in the previous two sections<sup>4</sup>. The experiment involves the simulation of samples of the CIR and OU processes and the comparison of the log-likelihood of the resultant samples

---

<sup>4</sup>A spectral method was also considered in this comparative exercise but because of space considerations the method is not described here nor the results reported. It turned out that the spectral method was superior to the finite-difference method, but inferior to the finite-element method. All the C code to implement all the comparative experiments (including those based on the spectral method) and the subsequent parameter estimation exercises are available from the corresponding author on request.

computed by the various numerical procedures with values obtained from the closed-form expressions for the CIR and OU transitional PDFs. In the experiment each of the numerical procedures will be implemented using both the transitional PDF approach and the transitional CDF specification. In order to ensure a fair comparison between the PDF- and CDF-based specifications, the initial condition for the CDF-based specification will use equivalent starting information and require the aid of the CDF of the Normal distribution.

Data are obtained by generating 2000 samples of 500 observations from the CIR model

$$dX = \alpha(\beta - X) dt + \sigma\sqrt{X} dW$$

with true parameters  $\alpha = 0.20$ ,  $\beta = 0.08$  and  $\sigma = 0.10$ , and from the OU process

$$dX = \alpha(\beta - X) dt + \sigma dW$$

with  $\alpha = 0.20$ ,  $\beta = 0.08$  and  $\sigma = 0.03$ . The synthetic samples are generated using the Milstein scheme with each time interval between observations  $\Delta = 1/12$  (representing monthly data) broken into 1000 steps to ensure that the observed data are accurate realisations of the process.

The log-likelihood of each sample is computed using the following approaches:

1. **Exact Likelihood** - based on the known closed form expressions for the transitional PDFs of the CIR and OU processes.
2. **Finite-difference procedures** - the PDF- and CDF-based procedures are both initialised at time  $t_k + \Delta_t$  using the Jensen and Poulsen (JP) approximation of the transitional PDF (the initialisation of the CDF utilises the Normal CDF). The procedures are run at all combinations of  $\Delta_x = 0.005, 0.002, 0.001$  and  $\Delta_t = 1/120, 1/600$ .
3. **Finite-element approach** - the PDF- and CDF-based procedures are both initialised at time  $t_k + \Delta_t$  using the Jensen and Poulsen (JP) approximation of the transitional PDF (the initialisation of the CDF utilises the Normal CDF). The procedures are run at all combinations of  $n = 50, 100, 200$  and  $\Delta_t = 1/120, 1/600$ .

The maximum absolute relative error, mean absolute relative error and the mean squared relative error between the various numerical estimates of log-likelihood and the exact log-likelihood over 2000 repetitions are presented in Tables 1 and 2. Table 1 presents the results for the finite-difference procedures and Table 2 presents the results for the finite-element procedures.



		Measures of relative error		
Finite difference	$\Delta_x$	Max. Absolute	Mean Absolute	Mean Squared
PDF-based procedure				
$\Delta_t = 1/120$	0.005	$1.132 \times 10^{-2}$	$2.561 \times 10^{-3}$	$1.036 \times 10^{-5}$
	0.002	$2.737 \times 10^{-3}$	$2.866 \times 10^{-4}$	$1.751 \times 10^{-7}$
	0.001	$1.236 \times 10^{-3}$	$6.123 \times 10^{-5}$	$8.989 \times 10^{-9}$
$\Delta_t = 1/600$	0.005	$1.744 \times 10^{-1}$	$9.211 \times 10^{-3}$	$1.624 \times 10^{-4}$
	0.002	$1.236 \times 10^{-2}$	$2.764 \times 10^{-3}$	$1.143 \times 10^{-5}$
	0.001	$3.825 \times 10^{-3}$	$3.189 \times 10^{-4}$	$2.529 \times 10^{-7}$
CDF-based procedure				
$\Delta_t = 1/120$	0.005	$2.153 \times 10^{-2}$	$4.439 \times 10^{-3}$	$2.715 \times 10^{-5}$
	0.002	$2.196 \times 10^{-3}$	$3.933 \times 10^{-4}$	$2.510 \times 10^{-7}$
	0.001	$1.043 \times 10^{-3}$	$1.036 \times 10^{-4}$	$2.003 \times 10^{-8}$
$\Delta_t = 1/600$	0.005	$2.373 \times 10^{-2}$	$5.007 \times 10^{-3}$	$3.398 \times 10^{-5}$
	0.002	$2.217 \times 10^{-3}$	$4.115 \times 10^{-4}$	$2.753 \times 10^{-7}$
	0.001	$1.135 \times 10^{-3}$	$1.017 \times 10^{-4}$	$1.932 \times 10^{-8}$
Finite element	$n$			
PDF-based procedure				
$\Delta_t = 1/120$	50	$1.847 \times 10^{-2}$	$6.636 \times 10^{-4}$	$1.591 \times 10^{-6}$
	100	$5.921 \times 10^{-4}$	$1.510 \times 10^{-4}$	$2.604 \times 10^{-8}$
	200	$2.860 \times 10^{-4}$	$3.953 \times 10^{-5}$	$2.453 \times 10^{-9}$
$\Delta_t = 1/600$	50	$1.706 \times 10^{-2}$	$1.346 \times 10^{-3}$	$3.292 \times 10^{-6}$
	100	$1.036 \times 10^{-3}$	$3.048 \times 10^{-4}$	$9.766 \times 10^{-8}$
	200	$2.207 \times 10^{-4}$	$7.598 \times 10^{-5}$	$6.048 \times 10^{-9}$
CDF-based procedure				
$\Delta_t = 1/120$	50	$1.916 \times 10^{-2}$	$2.084 \times 10^{-4}$	$1.487 \times 10^{-6}$
	100	$4.557 \times 10^{-4}$	$3.803 \times 10^{-5}$	$2.545 \times 10^{-9}$
	200	$2.463 \times 10^{-4}$	$2.412 \times 10^{-5}$	$1.043 \times 10^{-9}$
$\Delta_t = 1/600$	50	$1.916 \times 10^{-2}$	$2.315 \times 10^{-4}$	$1.768 \times 10^{-6}$
	100	$4.396 \times 10^{-4}$	$3.057 \times 10^{-5}$	$2.012 \times 10^{-9}$
	200	$2.222 \times 10^{-4}$	$7.320 \times 10^{-6}$	$1.407 \times 10^{-10}$

Table 1: Measures of relative error in the calculation of log-likelihood for the CIR process using the finite-difference and finite-element procedures and based on 2000 simulations.

		Measures of relative error		
<b>Finite difference</b>	$\Delta_x$	Max. Absolute	Mean Absolute	Mean Squared
PDF-based procedure				
$\Delta_t = 1/120$	0.005	$6.604 \times 10^{-3}$	$1.608 \times 10^{-3}$	$4.007 \times 10^{-6}$
	0.002	$7.342 \times 10^{-4}$	$1.155 \times 10^{-4}$	$2.194 \times 10^{-8}$
	0.001	$3.216 \times 10^{-4}$	$3.377 \times 10^{-5}$	$2.288 \times 10^{-9}$
$\Delta_t = 1/600$	0.005	$2.822 \times 10^{-2}$	$1.103 \times 10^{-2}$	$1.408 \times 10^{-4}$
	0.002	$4.270 \times 10^{-3}$	$1.046 \times 10^{-3}$	$1.667 \times 10^{-6}$
	0.001	$1.902 \times 10^{-4}$	$3.277 \times 10^{-5}$	$1.790 \times 10^{-9}$
CDF-based procedure				
$\Delta_t = 1/120$	0.005	$4.868 \times 10^{-3}$	$1.819 \times 10^{-3}$	$4.448 \times 10^{-6}$
	0.002	$1.234 \times 10^{-3}$	$2.175 \times 10^{-4}$	$7.354 \times 10^{-8}$
	0.001	$4.618 \times 10^{-4}$	$5.850 \times 10^{-5}$	$5.748 \times 10^{-9}$
$\Delta_t = 1/600$	0.005	$5.548 \times 10^{-3}$	$2.039 \times 10^{-3}$	$5.541 \times 10^{-6}$
	0.002	$1.156 \times 10^{-3}$	$2.219 \times 10^{-4}$	$7.603 \times 10^{-8}$
	0.001	$3.339 \times 10^{-4}$	$5.713 \times 10^{-5}$	$5.174 \times 10^{-9}$
<b>Finite element</b>	$n$			
PDF-based procedure				
$\Delta_t = 1/120$	50	$7.220 \times 10^{-4}$	$4.060 \times 10^{-4}$	$1.730 \times 10^{-7}$
	100	$3.068 \times 10^{-4}$	$1.024 \times 10^{-4}$	$1.184 \times 10^{-8}$
	200	$2.555 \times 10^{-4}$	$3.105 \times 10^{-5}$	$1.655 \times 10^{-9}$
$\Delta_t = 1/600$	50	$1.065 \times 10^{-3}$	$7.792 \times 10^{-4}$	$6.161 \times 10^{-7}$
	100	$2.791 \times 10^{-4}$	$2.017 \times 10^{-4}$	$4.110 \times 10^{-8}$
	200	$7.168 \times 10^{-5}$	$4.997 \times 10^{-5}$	$2.526 \times 10^{-9}$
CDF-based procedure				
$\Delta_t = 1/120$	50	$6.499 \times 10^{-4}$	$7.032 \times 10^{-5}$	$8.617 \times 10^{-9}$
	100	$1.785 \times 10^{-4}$	$2.673 \times 10^{-5}$	$1.166 \times 10^{-9}$
	200	$2.238 \times 10^{-4}$	$2.267 \times 10^{-5}$	$9.315 \times 10^{-10}$
$\Delta_t = 1/600$	50	$9.467 \times 10^{-4}$	$7.375 \times 10^{-5}$	$1.171 \times 10^{-8}$
	100	$1.843 \times 10^{-4}$	$1.628 \times 10^{-5}$	$4.988 \times 10^{-10}$
	200	$4.019 \times 10^{-5}$	$4.098 \times 10^{-6}$	$2.953 \times 10^{-11}$

Table 2: Measures of relative error in the calculation of log-likelihood for the OU process using the finite-difference and finite-element procedures and based on 2000 simulations.

A number of important general findings are apparent from these tables.

1. All of the numerical procedures deliver accurate estimates of log-likelihood.
2. The PDF-based procedures are often less accurate than would be anticipated at fine temporal resolutions, because when  $\Delta_t$  is small the JP initial condition may be distinguishable from zero at only a few (or even zero) nodes. This highlights a possible problem with the PDF specifications in parameter estimation applications. If the optimisation routine proposes parameter values that cause the initial condition to be not very diffuse relative to the chosen spatial discretisation, the resulting estimate of log-likelihood will be inaccurate and in extreme cases may even cause the optimisation procedure to fail.
3. The CDF-based procedures are generally as accurate as, and more robust than the equivalent PDF-based procedures.
4. The finite-element procedures are able to return better accuracy than the finite-difference procedures even when using fewer spatial nodes. It is difficult to directly compare the number of nodes used, because the finite-difference procedures require an evenly spaced grid across  $[x_0, x_n]$  and thus use a different number of nodes for each sample. However, the average size of the interval  $[x_0, x_n]$  over the 2000 samples (for both the OU and CIR processes) is approximately 0.4. This implies that a spatial discretisation of  $\Delta_x = 0.005$  in the finite-difference procedure equates to roughly 80 nodes,  $\Delta_x = 0.002$  equates to roughly 200 nodes, and  $\Delta_x = 0.001$  equates to roughly 400 nodes.

## 6 Interpolation

The SDE models commonly used in finance are almost invariably independent of time, that is, the specifications of drift and diffusion are functions of state alone. This means that transitions from the same state at different times can be treated using the same transitional PDF, and moreover, the transitional PDFs for similar initial values can be expected to exhibit quantitatively similar behaviour. This observation suggests that all the transitions within a sample, no matter how large the sample, may be characterised by a reduced set of transitional PDFs which, when combined in an appropriate way, enable the likelihood of every transition within a sample to be computed to high accuracy.

In practice, each member of this reduced set of transitional PDFs will be defined by its initial state, and the set of all such states will be denoted by  $\mathcal{B}$ . The likelihood for any particular transition in the sample will be computed by interpolation of the transitional densities with initial states  $\mathcal{B}$ . The procedure is implemented as follows.

1. Choose a discrete set of initial states, say  $\mathcal{B} = \{b_0, b_1, \dots, b_M\}$  so that the observations in the

sample lie within the support of  $\mathcal{B}$ . It is envisaged that the number of initial states  $(M + 1)$  will be substantially less than the number of observations in the sample  $(N + 1)$ .

2. Any of the procedures described in Sections 3 and 4 can be used to provide the likelihoods of transitions from the initial state  $b_j$  to the observations  $X_1, \dots, X_N$  in the sample. When performed over all the states in  $\mathcal{B}$ , the result of these computations may be encapsulated in a matrix  $\Psi$  of dimension  $N \times (M + 1)$  in which  $\Psi_{ij}$  is the likelihood of a transition from the initial state  $b_j$  to the state  $X_{i+1}$ . Thus each column of  $\Psi$  will correspond to the likelihood of a transition from the initial state associated with the column to each state of the sample.
3. Without loss of generality, the particular transition in the observed data for which the transitional density is required may be taken to be that from  $X_0$  to  $X_1$ . The first row of  $\Psi$  contains the likelihoods of a transition from  $\mathcal{B}$  to state  $X_1$ . Of course,  $X_0$  itself is most unlikely to be an element of  $\mathcal{B}$ , and so the transitional density from  $X_0$  to  $X_1$  is calculated in general by interpolating the first row of the matrix  $\Psi$ .
4. This interpolation procedure is repeated for each transition from  $X_k$  to  $X_{k+1}$  for integers  $k$  between  $k = 1$  and  $k = N - 1$  inclusive. The resulting likelihoods are then used to compute the total negative log-likelihood of the sample.

The mechanics of the algorithm is illustrated with reference to a sample of eleven observations in the interval  $[0.02, 0.10]$ . The set  $\mathcal{B} = \{0.02, 0.04, 0.06, 0.08, 0.10\}$  and the rows in Table 3 contain the likelihoods of making the transition from each initial state to the relevant observation. For example, the first entry 6.857 is the likelihood of a transition from  $b_0 = 0.02$  to  $X_1 = 0.054$ , that is, the value of the transitional density  $f(X_1 = 0.054 | b_0 = 0.02)$ .

Observations	Initial States				
	$b_0 = 0.02$	$b_1 = 0.04$	$b_2 = 0.06$	$b_3 = 0.08$	$b_4 = 0.10$
$X_0 = 0.050$					
$X_1 = 0.054$	6.857	20.972	20.437	10.411	3.432
$X_2 = 0.035$	28.855	21.264	6.870	1.374	0.198
$X_3 = 0.056$	5.562	19.435	20.996	11.670	4.155
$X_4 = 0.049$	11.740	24.369	17.590	6.951	1.831
$X_5 = 0.032$	31.648	18.310	4.900	0.836	0.105
$X_6 = 0.048$	12.656	24.710	17.010	6.457	1.642
$X_7 = 0.051$	9.660	23.289	18.877	8.229	2.364
$X_8 = 0.062$	2.844	14.576	21.178	15.123	6.717
$X_9 = 0.053$	7.420	21.535	20.152	9.919	3.173
$X_{10} = 0.092$	0.044	1.223	6.608	14.319	16.934

Table 3: Likelihoods of the transitions from a set  $\mathcal{B}$  of five initial states to each observation of the sample are given.

Consider, for example, the transition in the sample from  $x = X_0$  to  $x = X_1$ . The traditional approach would require, first, the distribution of transitional density to be obtained by solving the Fokker-Planck equation with initial state  $X_0$ , and then second, the evaluation of this solution at  $X_1$  to get the likelihood of the transition from  $X_0$  to  $X_1$ . To evaluate the likelihood function for the sample would require this procedure to be repeated ten times involving the computation of ten different solutions of the Fokker-Planck equation.

By contrast, the interpolation procedure begins by solving the Fokker-Planck equation five times and uses these solutions to complete the columns of Table 3. The likelihood of each transition is then obtained by interpolating the appropriate row in the table. For example, the first transition in the data satisfies  $b_1 = 0.04 < X_0 < 0.06 = b_2$ , and therefore the likelihood of this transition may be estimated (but not necessarily in an optimal way) by linear interpolation between the densities computed for the initial states  $b_1$  and  $b_2$  to obtain

$$\text{Likelihood} = \frac{20.972(0.06 - 0.05) + 20.437(0.05 - 0.04)}{0.06 - 0.04} = 20.705.$$

To evaluate the likelihood function for the entire sample would therefore require the Fokker-Planck equation to be solved five times and these solutions to be interpolated ten times. The computational benefits of this procedure stem from the recognition that interpolation is a much cheaper operation than the numerical solution of a partial differential equation.

The crucial issues in the implementation of this approach are the speed and accuracy of the interpolation procedure. Linear interpolation was used by way of illustration in the example of Table 3 because of its simplicity. However, there are many possible ways to interpolate the densities. Two of which are discussed in the following subsection.

## 6.1 Interpolation Methods

Two interpolation methods, namely, Chebyshev Economisation and spline interpolation are now described.

### 6.1.1 Chebyshev Economisation

The first, and as it turns out, crucial step in providing a general interpolating procedure is the choice of the nodes  $\mathcal{B} = \{b_0, b_1, \dots, b_M\}$ . Intuitively, it would seem that the best way to place these nodes would be to spread them out equally across  $[a, b]$  including both endpoints. However, it is well known that this choice causes unacceptable oscillations in the interpolating polynomial  $p_M(x)$  near the ends of the interval of interpolation. Furthermore, this problem can not be alleviated simply by increasing the cardinality of  $\mathcal{B}$  as this just causes the problem to migrate closer to the endpoints of the interval and the oscillations to become more severe. The optimal choice of nodes is derived from the following important result. Suppose that the arbitrary function  $f(z)$  is to be approximated in

$[-1, 1]$  by a polynomial  $p_M(z)$  of degree  $M$  based on the nodes  $(z_0, f_0), \dots, (z_M, f_M)$  where, without loss of generality,  $z_0, \dots, z_M$  may be taken to be a strictly increasing sequence in  $[-1, 1]$ , then there exists  $\xi(z) \in (-1, 1)$  such that  $f(z)$  and  $p_M(z)$  satisfy

$$f(z) - p_M(z) = \frac{f^{(M+1)}(\xi(z))}{(M+1)!} \prod_{j=0}^M (z - z_j). \quad (44)$$

The interpolation error given by the right hand side of equation (44) consists of two parts; one part determined by the properties of the function and therefore beyond the user's control, and a second part determined by the choice of the nodes  $z_0, \dots, z_M$  and so entirely within the user's control. Chebyshev Economisation takes advantage of the optimal result

$$\text{Max} \left| \prod_{j=0}^M (z - z_j) \right| \geq \frac{1}{2^M}, \quad z, z_j \in [-1, 1], \quad (45)$$

in which equality is attained if and only if the nodes  $z_0, \dots, z_M$  are the zeros of the Chebyshev polynomial  $T_{M+1}(z)$  of order  $M+1$  which is defined on the interval  $[-1, 1]$  by

$$T_{M+1}(\cos \psi) = \cos [(M+1)\psi], \quad (46)$$

and has zeros at the points

$$z_j = \cos \left[ \frac{(2j+1)\pi}{2(M+1)} \right] \quad j = 0, \dots, M. \quad (47)$$

These nodes minimise the contribution made by the choice of nodes to the overall interpolation error, and equally importantly in the use of the interpolation procedure in parameter estimation, ensure that interpolation error is spread uniformly over  $[-1, 1]$ .

This result establishes the optimal choice for the nodes in the interval  $[-1, 1]$ . In general, however, the sample space of the problem may be taken to be  $[a, b]$ . This is easily dealt with by using the mapping

$$x = a + \frac{(b-a)(1+z)}{2}, \quad z \in [-1, 1]. \quad (48)$$

Consequently, the initial states  $b_0, b_1, \dots, b_M$  comprising  $\mathcal{B}$  are formed by mapping the zeros of the Chebyshev polynomial  $T_{M+1}(z)$  into  $[a, b]$  using (48) and (47) to obtain

$$b_j = a + (b-a) \cos^2 \left[ \frac{(2j+1)\pi}{4(M+1)} \right], \quad j = 0, \dots, M. \quad (49)$$

The uniform nature of the interpolation error in the Chebyshev Economisation makes it ideally suited to the calculation of a likelihood function involving transitions from initial states that are distributed over the entire interval  $[a, b]$ .

With the choice of nodes described in equation (49) an interpolated value of the target function at a particular point  $X_k$  can be obtained directly from the well known Lagrange interpolating polynomial

$$p_M(x) = \sum_{j=0}^M f_j L_{M,j}(z) \quad (50)$$

where  $f_j = f(z_j)$ ,  $z$  is obtained from  $x$  via the mapping in equation (48),  $L_{M,j}(z)$  is defined by

$$L_{M,j}(z) = \frac{(z - z_0)(z - z_1) \cdots (z - z_{j-1})(z - z_{j+1}) \cdots (z - z_M)}{(z_j - z_0)(z_j - z_1) \cdots (z_j - z_{j-1})(z_j - z_{j+1}) \cdots (z_j - z_M)}, \quad (51)$$

and, as before,  $z_0, z_1, \dots, z_M$  are the zeros of  $T_{M+1}(z)$ .

It transpires that the interpolation equation (50) may be efficiently computed by observing that  $L_{M,j}(z)$  based on the zeros of  $T_{M+1}(x)$  can be expressed in the compact form

$$L_{M,j}(z) = \frac{T_{M+1}(z)}{(z - z_j)T'_{M+1}(z_j)}. \quad (52)$$

Given  $z \in [-1, 1]$ , the defining equation (46) can be used to compute  $T_{M+1}(z)$  without recourse to repeated multiplication of differences of  $z$ . Further computational efficiencies may be realised by recognising, from (46) that

$$T'_{M+1}(z_j) = \frac{(-1)^j(M+1)}{(1 - z_j^2)^{1/2}}. \quad (53)$$

### 6.1.2 Spline interpolation

By contrast with Chebyshev Economisation, spline interpolation is not based on a single interpolating polynomial, but rather on a piecewise function, known as a spline function, that is constructed from a sequence of lower order polynomials. The benefit of this approach is that the low order of the polynomials from which the spline function is constructed reduces its propensity to oscillate near the endpoints of the interpolating interval, even if the interpolating nodes are equally spaced.

Spline interpolation involves breaking the state space  $[a, b]$  into  $M$  subintervals

$$a = b_0 < b_1 < \dots < b_M = b$$

where the points  $b_0, \dots, b_M$  are referred to as knots<sup>5</sup> and are typically evenly spaced (but do not have to be). For each subinterval, a low-order polynomial is used to approximate the target function and a spline function is constructed by joining together this sequence of low-order polynomials. To ensure that the spline function has some degree of smoothness the polynomial for each subinterval is usually chosen such that the resulting spline function and its first few derivatives are continuous, particularly at the internal knots.

The most widely used spline functions are cubic splines, which are constructed by joining together  $M$  cubic polynomials, and the remainder of this subsection will focus on this case. The cubic polynomial for the  $j$ -th subinterval is defined by

$$s_j(x) = c_{0,j} + c_{1,j}(x - b_j) + c_{2,j}(x - b_j)^2 + c_{3,j}(x - b_j)^3 \quad (54)$$

---

<sup>5</sup>Note that it is not necessary for the knots and the nodes in  $\mathcal{B}$  to be identical, but, for convenience that is the convention adopted here.

where  $c_{0,j}, c_{1,j}, c_{2,j}$  and  $c_{3,j}$  are parameters to be estimated and  $j = 0, \dots, M$ . The cubic polynomial for each subinterval contains four unknown parameters and so there is a total of  $4M$  parameters to be estimated. The condition that the polynomial for each subinterval must equal the known function values at the knots at the start and end of the subinterval provides  $2M$  equations and conditions requiring continuity of first and second derivatives at internal knots provide an extra  $2(M - 1)$  equations. Hence, two more equations are needed to identify the parameters. These are often supplied by a condition known as the “not-a-knot” condition, which involves forcing the third derivative of the spline function to be continuous at the first and last internal knots. With as many equations as unknown parameters, the parameters can be easily identified using Gaussian elimination. Once the parameters are known, the interpolated value of the function at any point  $X_k$  can be found by identifying which subinterval of the spline  $X_k$  is in and then evaluating the cubic polynomial for that subinterval at  $X_k$ .

## 6.2 Interpolation error

This section outlines an experiment designed to investigate how the magnitude of interpolation error depends on the dimension of  $\mathcal{B}$ , the set of initial states to be used by the interpolation procedure. In overview, three distributions with closed-form expressions for their PDFs are used to assign densities at the nodes  $\mathcal{B}$  contained within the interval  $[a, b]$  of state space. For each distribution, these limits are set to ensure that 99.998% of the probability mass lies within the interval. For each density, the experiment is repeated for  $M = 10, 15, 20, 25$  and 30 nodes in  $\mathcal{B}$ . In any experiment, the values of the densities at the nodes in  $\mathcal{B}$  are used to estimate the PDF over a mesh of 10000 uniformly spaced points spanning  $[a, b]$  based on both Chebyshev Economisation and a cubic spline with not-a-knot end conditions. These estimates are compared with the true values computed from the closed-form expression for the transitional PDF, and the size of the relative error resulting from the interpolation procedure is measured using the integrated  $L^1$  and  $L^2$  norms. The model PDFs used in the experiment with their chosen parameter values are given in Table 4.

Distribution	Probability Density Function	Parameter Values
Gaussian	$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$	$\mu = 6.062 \times 10^{-2}$ $\sigma = 1.863 \times 10^{-2}$
Log-normal	$f(x; \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$	$\mu = -2.853$ $\sigma = 3.255 \times 10^{-1}$
Gamma	$f(x; \alpha, \lambda) = \frac{1}{\lambda \Gamma(\alpha)} \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left[-\frac{x}{\lambda}\right]$	$\alpha = 1.012 \times 10^1$ $\lambda = 5.992 \times 10^{-3}$

Table 4: Model distributions used in the investigation of the magnitude of interpolation error

The choice of PDFs is motivated by the frequency with which the underlying distributions arise in



economic and financial applications, and in particular as transitional PDFs of processes specified by SDEs. For example, the transitional PDF associated with Brownian motion is the Gaussian distribution or the Lognormal distribution for the case of geometric Brownian motion, while the steady state transitional density of a CIR process is a Gamma distribution. To ground the experiment in a realistic setting, the parameter values in Table 4 were obtained by fitting each density to interest rate data as though it represented the transitional PDF of a SDE modelling short-term interest rates. Table 5 reports the results of this experiment.

Distribution	$M$	Norms of relative error using cubic spline interpolation		Norms of relative error using Chebyshev Economisation	
		$L^1$	$L^2$	$L^1$	$L^2$
<b>Gaussian</b>	10	$3.508 \times 10^{-1}$	5.620	$6.738 \times 10^{-1}$	$2.363 \times 10^1$
	15	$2.622 \times 10^{-2}$	$4.146 \times 10^{-2}$	$5.587 \times 10^{-2}$	$1.925 \times 10^{-1}$
	20	$5.039 \times 10^{-3}$	$1.844 \times 10^{-3}$	$1.290 \times 10^{-3}$	$9.638 \times 10^{-5}$
	25	$1.465 \times 10^{-3}$	$1.789 \times 10^{-4}$	$1.914 \times 10^{-5}$	$2.350 \times 10^{-8}$
	30	$5.480 \times 10^{-4}$	$2.789 \times 10^{-5}$	$2.108 \times 10^{-7}$	$3.000 \times 10^{-12}$
<b>Log-normal</b>	10	$2.107 \times 10^{-1}$	2.785	3.607	$4.385 \times 10^2$
	15	$1.125 \times 10^{-1}$	2.234	$1.070 \times 10^{-1}$	$4.073 \times 10^{-1}$
	20	$7.441 \times 10^{-2}$	1.057	$2.109 \times 10^{-2}$	$1.359 \times 10^{-2}$
	25	$2.932 \times 10^{-2}$	$1.801 \times 10^{-1}$	$3.517 \times 10^{-4}$	$1.617 \times 10^{-5}$
	30	$9.462 \times 10^{-3}$	$2.019 \times 10^{-2}$	$1.734 \times 10^{-4}$	$1.025 \times 10^{-6}$
<b>Gamma</b>	10	$1.889 \times 10^{-1}$	4.383	1.425	$1.197 \times 10^2$
	15	$5.139 \times 10^{-2}$	$4.139 \times 10^{-1}$	$1.670 \times 10^{-2}$	$1.542 \times 10^{-2}$
	20	$7.022 \times 10^{-3}$	$8.378 \times 10^{-3}$	$1.514 \times 10^{-3}$	$1.320 \times 10^{-4}$
	25	$4.973 \times 10^{-4}$	$5.131 \times 10^{-5}$	$2.182 \times 10^{-5}$	$2.864 \times 10^{-8}$
	30	$1.081 \times 10^{-3}$	$3.079 \times 10^{-4}$	$2.004 \times 10^{-7}$	$2.000 \times 10^{-12}$

Table 5:  $L^1$  and  $L^2$  norms of the relative error in estimated density using cubic spline interpolation and Chebyshev Economisation with varying numbers of nodes.

The results indicate that Chebyshev Economisation is very accurate and dominates cubic spline interpolation once the number of nodes exceeds 20. However, both the construction of this experiment and the use of closed-form expressions for the PDFs ensure that all densities can be computed to high accuracy. By contrast, typical transitional densities are computed by means of a numerical procedure, are often strongly localised within  $[a, b]$  and are less accurate than would be obtained from a closed-form expression. In practical applications, therefore, the value of the density at many nodes may be poorly resolved and indistinguishable from zero. As a consequence, the actual dimension of  $\mathcal{B}$  in

practice must often be larger than the results of the experiment reported in Table 5 would suggest<sup>6</sup>.

### 6.3 Log-likelihood computation

This section outlines a simulation experiment designed to explore the efficacy of the interpolation procedure in the context of likelihood computation. The experiment involves the simulation of samples of the CIR and OU models followed by the computation of the log-likelihoods of the samples by means of Chebyshev Economisation, and the comparison of these log-likelihoods with their true values. In addition to the 2000 samples of CIR and OU data containing  $N = 500$  transitions generated in Section 5 a further 2000 samples are generated using an identical procedure but each containing  $N = 1000$  transitions.

For each sample, the upper and lower limits of the interpolating interval  $[x_0, x_n]$  are respectively the maximum and minimum observations in the sample. For each process the known closed-form expression for the transitional PDF is used to construct the  $N \times (M + 1)$  matrix,  $\Psi$ , in which each column corresponds to the likelihood of a transition from a particular initial state to all states in the sample. This procedure is implemented for the cases  $M = 50$ ,  $M = 75$  and  $M = 100$ . Chebyshev Economisation is used to estimate the log-likelihood of the sample which is then compared with that calculated from the closed-form expression for the transitional PDF. The maximum absolute relative error, mean absolute relative error and the mean squared relative error are calculated from the 2000 repetitions. The results of these simulation exercises are presented in Table 6.

The accuracy of the Chebyshev Economisation in this controlled environment is quite remarkable, particularly with the choices  $M = 75$  and  $M = 100$ . The relatively poorer results with  $M = 50$  are dominated by a few simulations. In the overwhelming majority of simulations the log-likelihood is computed to very high accuracy, but in a few simulations (less than 0.5% of the total simulations) a small negative value is proposed for the likelihood of a transition, which incurs a substantial penalty. Note that as a closed-form expression for the transitional PDF is being used to generate the values of the function required to perform the interpolation, the effect of sample size is simply to increase the scope for error. The slight increase in the error norms for the higher sample sizes is therefore expected.

---

<sup>6</sup>In the empirical application presented in Section 8 of this paper a quick calibration exercise is performed to ensure that a large enough number of interpolating nodes are employed.

	$M$	Measures of relative error		
		Maximum Absolute	Mean Absolute	Mean Squared
CIR Process				
$N = 500$	50	$3.965 \times 10^{-2}$	$5.472 \times 10^{-5}$	$1.287 \times 10^{-6}$
	75	$4.312 \times 10^{-6}$	$1.231 \times 10^{-8}$	$2.002 \times 10^{-14}$
	100	$2.964 \times 10^{-8}$	$1.823 \times 10^{-11}$	$4.418 \times 10^{-19}$
$N = 1000$	50	$9.802 \times 10^{-3}$	$1.043 \times 10^{-4}$	$7.271 \times 10^{-7}$
	75	$9.112 \times 10^{-5}$	$1.119 \times 10^{-7}$	$4.511 \times 10^{-12}$
	100	$6.404 \times 10^{-8}$	$8.361 \times 10^{-11}$	$2.913 \times 10^{-18}$
OU Process				
$N = 500$	50	$1.878 \times 10^{-2}$	$2.877 \times 10^{-5}$	$3.228 \times 10^{-7}$
	75	$1.711 \times 10^{-5}$	$3.103 \times 10^{-8}$	$2.147 \times 10^{-13}$
	100	$3.179 \times 10^{-9}$	$1.185 \times 10^{-11}$	$2.183 \times 10^{-20}$
$N = 1000$	50	$1.879 \times 10^{-2}$	$1.723 \times 10^{-4}$	$1.410 \times 10^{-6}$
	75	$2.237 \times 10^{-5}$	$1.370 \times 10^{-7}$	$1.156 \times 10^{-12}$
	100	$5.029 \times 10^{-8}$	$1.613 \times 10^{-10}$	$3.108 \times 10^{-18}$

Table 6: Measures of relative error in the calculation of log-likelihood for the CIR and OU processes using Chebyshev Economisation and based on 2000 simulations.

## 7 Parameter estimation

In the experiment in Section 6.3, the transitional densities are obtained from closed-form expressions. In practice, however, such expressions rarely exist, and consequently the densities to be interpolated are obtained by other methods, such as the finite-difference and finite-element procedures outlined earlier. In the experiment conducted in this section, the parameters of the CIR and OU models are estimated for the same data as that used in Section 6.3.

For the sake of comparison, the closed-form expression for the transitional PDF will be used to provide EML estimates of the parameters. These will then be compared to parameter estimates based on the four procedures for solving the Fokker-Planck equation numerically introduced in Sections 3 and 4. All four procedures are employed with and without interpolation. All of the numerical procedures use a temporal resolution  $\Delta_t = 1/120$ . The finite-difference approaches are run with  $\Delta_x = 0.001$  and the finite-element approaches are run with  $n = 100$ . In all of the procedures using the interpolation technique 76 interpolating nodes ( $M = 75$ ) are used.

Tables 7 and 8 report the average difference between the EML parameter estimates and the estimates from the numerical procedures over the 2000 samples for the CIR and OU processes respectively. The standard deviation of these differences and the average amount of computational time taken per parameter estimation are also reported.

There are a number of important conclusions that emerge from these results. The first general point to note about these results is that *all* of the numerical procedures (for all practical purposes) return the EML parameter estimates not only on average, but in each and every sample. Turning now to more specific points relating to the methodological innovations suggested in this paper, the following three conclusions may be drawn.

1. The CDF-based procedures are slightly slower than the equivalent PDF-based procedures, but as noted earlier in the experiment in Section 6.3 they are undoubtedly more robust.
2. In terms of computational efficiency the finite-element procedures far outperform the finite-difference procedures.
3. The computational burden of all of the procedures is reduced significantly when used in combination with the interpolation technique. Moreover, the average parameter estimates for the procedures making use of the interpolation technique are virtually indistinguishable from the estimates provided by equivalent procedure without interpolation.

It is perhaps worth emphasizing some issues relating to the efficacy of the interpolation idea. Of particular importance is the fact that for the procedures using the interpolation idea the mean times taken to perform the estimation for the sample size  $N = 1000$  are only marginally greater than the mean time taken for a sample size  $N = 500$ . This is in startling contrast to the time taken by the procedures without interpolation for which the average computational time roughly doubles as the sample size is doubled. As pointed out earlier, the efficiency of the procedures using the interpolation technique is due to the fact that, irrespective of the sample size, the Fokker-Planck equation need only be solved as many times as there are interpolating nodes (76 in this case). Any increase in computing time experienced for larger samples is then entirely due to the increased interpolation and, as shown by this example, this tends to be trivial.

In summary, this experiment would suggest that the CDF-based finite-element procedure in combination with the interpolation technique is a very accurate, computationally efficient and robust estimation procedure which could stand comparison with any other approach to estimating the parameters of SDEs.

Procedure	Specification	Mean difference between EML estimate and numerical estimate			Time (sec)
		$\alpha$	$\beta$	$\sigma$	
N=500					
Finite Diff.	PDF	$5.083 \times 10^{-4}$ ( $7.672 \times 10^{-3}$ )	$9.939 \times 10^{-5}$ ( $2.530 \times 10^{-3}$ )	$-1.871 \times 10^{-4}$ ( $1.307 \times 10^{-4}$ )	8.13
	PDF (interp.)	$5.940 \times 10^{-4}$ ( $6.980 \times 10^{-3}$ )	$8.686 \times 10^{-5}$ ( $2.464 \times 10^{-3}$ )	$-1.821 \times 10^{-4}$ ( $2.407 \times 10^{-4}$ )	1.37
	CDF	$9.787 \times 10^{-4}$ ( $8.212 \times 10^{-3}$ )	$1.234 \times 10^{-4}$ ( $2.540 \times 10^{-3}$ )	$-5.329 \times 10^{-4}$ ( $1.974 \times 10^{-4}$ )	8.63
	CDF (interp.)	$1.020 \times 10^{-3}$ ( $8.115 \times 10^{-3}$ )	$1.082 \times 10^{-4}$ ( $2.579 \times 10^{-3}$ )	$-5.331 \times 10^{-4}$ ( $1.990 \times 10^{-4}$ )	1.48
Finite Element	PDF	$4.880 \times 10^{-4}$ ( $7.902 \times 10^{-3}$ )	$3.678 \times 10^{-6}$ ( $2.832 \times 10^{-3}$ )	$-1.100 \times 10^{-4}$ ( $8.016 \times 10^{-5}$ )	3.00
	PDF (interp.)	$3.045 \times 10^{-4}$ ( $6.030 \times 10^{-3}$ )	$7.360 \times 10^{-5}$ ( $2.520 \times 10^{-3}$ )	$-8.055 \times 10^{-5}$ ( $7.285 \times 10^{-5}$ )	0.90
	CDF	$1.899 \times 10^{-4}$ ( $9.877 \times 10^{-3}$ )	$6.258 \times 10^{-5}$ ( $2.456 \times 10^{-3}$ )	$-9.753 \times 10^{-5}$ ( $8.073 \times 10^{-5}$ )	3.80
	CDF (interp.)	$3.195 \times 10^{-4}$ ( $6.246 \times 10^{-3}$ )	$8.442 \times 10^{-5}$ ( $2.441 \times 10^{-3}$ )	$-7.226 \times 10^{-5}$ ( $7.595 \times 10^{-5}$ )	1.02
N=1000					
Finite Diff.	PDF	$6.152 \times 10^{-4}$ ( $6.766 \times 10^{-3}$ )	$3.280 \times 10^{-5}$ ( $1.786 \times 10^{-3}$ )	$-1.863 \times 10^{-4}$ ( $1.023 \times 10^{-4}$ )	17.13
	PDF (interp.)	$5.940 \times 10^{-4}$ ( $6.980 \times 10^{-3}$ )	$-3.292 \times 10^{-6}$ ( $1.838 \times 10^{-3}$ )	$-1.760 \times 10^{-4}$ ( $1.359 \times 10^{-3}$ )	1.76
	CDF	$1.274 \times 10^{-3}$ ( $9.306 \times 10^{-3}$ )	$1.438 \times 10^{-5}$ ( $1.818 \times 10^{-3}$ )	$-5.322 \times 10^{-4}$ ( $1.490 \times 10^{-4}$ )	18.35
	CDF (interp.)	$1.369 \times 10^{-3}$ ( $7.630 \times 10^{-3}$ )	$2.147 \times 10^{-5}$ ( $1.818 \times 10^{-3}$ )	$-5.320 \times 10^{-4}$ ( $1.489 \times 10^{-4}$ )	1.89
Finite Element	PDF	$5.326 \times 10^{-4}$ ( $6.322 \times 10^{-3}$ )	$4.087 \times 10^{-6}$ ( $1.791 \times 10^{-3}$ )	$-8.870 \times 10^{-5}$ ( $7.070 \times 10^{-5}$ )	5.57
	PDF (interp.)	$4.546 \times 10^{-4}$ ( $6.604 \times 10^{-3}$ )	$-3.938 \times 10^{-6}$ ( $1.790 \times 10^{-3}$ )	$-9.122 \times 10^{-5}$ ( $1.005 \times 10^{-4}$ )	1.36
	CDF	$4.496 \times 10^{-4}$ ( $7.762 \times 10^{-3}$ )	$-2.103 \times 10^{-5}$ ( $1.829 \times 10^{-3}$ )	$-7.754 \times 10^{-5}$ ( $6.660 \times 10^{-5}$ )	7.02
	CDF (interp.)	$2.793 \times 10^{-4}$ ( $7.352 \times 10^{-3}$ )	$2.705 \times 10^{-5}$ ( $1.866 \times 10^{-3}$ )	$-8.153 \times 10^{-5}$ ( $1.084 \times 10^{-4}$ )	1.47

Table 7: Mean difference between EML estimates of the parameters of the CIR model and the estimates obtained from a variety of numerical procedures. The average is taken over 2000 different samples of the process  $dX = \alpha(\beta - X)dt + \sigma\sqrt{X}dW$  with parameters  $\alpha = 0.2$ ,  $\beta = 0.08$  and  $\sigma = 0.10$ . The standard deviation of the difference between the estimates is shown in parentheses. The average EML parameter estimates over the 2000 samples of  $N = 500$  transitions are  $\alpha = 0.3097$ ,  $\beta = 0.0811$  and  $\sigma = 0.1002$ . The average EML parameter estimates over the 2000 samples of  $N = 1000$  transitions are  $\alpha = 0.2484$ ,  $\beta = 0.0807$  and  $\sigma = 0.1000$ .

Procedure	Specification	Mean difference between EML estimate and numerical estimate			Time (sec)
		$\alpha$	$\beta$	$\sigma$	
N=500					
Finite Diff.	PDF	$-5.136 \times 10^{-5}$ ( $5.826 \times 10^{-3}$ )	$7.939 \times 10^{-5}$ ( $3.052 \times 10^{-3}$ )	$-3.657 \times 10^{-5}$ ( $2.491 \times 10^{-5}$ )	8.11
	PDF (interp.)	$6.613 \times 10^{-5}$ ( $3.747 \times 10^{-3}$ )	$9.080 \times 10^{-4}$ ( $5.179 \times 10^{-4}$ )	$-3.637 \times 10^{-5}$ ( $2.272 \times 10^{-5}$ )	1.38
	CDF	$-2.894 \times 10^{-4}$ ( $8.806 \times 10^{-3}$ )	$1.033 \times 10^{-4}$ ( $3.028 \times 10^{-3}$ )	$-1.025 \times 10^{-4}$ ( $2.791 \times 10^{-5}$ )	8.86
	CDF (interp.)	$-1.445 \times 10^{-4}$ ( $5.584 \times 10^{-3}$ )	$2.928 \times 10^{-5}$ ( $6.171 \times 10^{-4}$ )	$-1.020 \times 10^{-4}$ ( $2.336 \times 10^{-5}$ )	1.50
Finite Element	PDF	$-4.661 \times 10^{-4}$ ( $1.196 \times 10^{-2}$ )	$6.945 \times 10^{-5}$ ( $3.046 \times 10^{-3}$ )	$-1.952 \times 10^{-5}$ ( $2.477 \times 10^{-5}$ )	3.04
	PDF (interp.)	$-2.212 \times 10^{-4}$ ( $5.170 \times 10^{-3}$ )	$6.664 \times 10^{-5}$ ( $3.087 \times 10^{-3}$ )	$-1.825 \times 10^{-5}$ ( $2.879 \times 10^{-5}$ )	0.86
	CDF	$-1.900 \times 10^{-4}$ ( $6.098 \times 10^{-3}$ )	$8.157 \times 10^{-5}$ ( $3.090 \times 10^{-3}$ )	$-1.632 \times 10^{-5}$ ( $3.079 \times 10^{-5}$ )	3.86
	CDF (interp.)	$-1.794 \times 10^{-4}$ ( $5.553 \times 10^{-3}$ )	$1.343 \times 10^{-5}$ ( $6.791 \times 10^{-4}$ )	$-1.595 \times 10^{-5}$ ( $2.916 \times 10^{-5}$ )	0.98
N=1000					
Finite Diff.	PDF	$9.130 \times 10^{-5}$ ( $7.550 \times 10^{-3}$ )	$-1.831 \times 10^{-5}$ ( $8.287 \times 10^{-4}$ )	$-3.468 \times 10^{-5}$ ( $2.688 \times 10^{-5}$ )	17.92
	PDF (interp.)	$6.330 \times 10^{-5}$ ( $7.441 \times 10^{-3}$ )	$9.017 \times 10^{-4}$ ( $1.003 \times 10^{-3}$ )	$-3.468 \times 10^{-5}$ ( $2.706 \times 10^{-5}$ )	1.75
	CDF	$2.496 \times 10^{-4}$ ( $6.335 \times 10^{-3}$ )	$-1.507 \times 10^{-5}$ ( $8.304 \times 10^{-4}$ )	$-1.004 \times 10^{-4}$ ( $2.589 \times 10^{-5}$ )	19.44
	CDF (interp.)	$2.254 \times 10^{-4}$ ( $6.289 \times 10^{-3}$ )	$-3.254 \times 10^{-6}$ ( $1.037 \times 10^{-3}$ )	$-1.001 \times 10^{-4}$ ( $2.742 \times 10^{-5}$ )	1.94
Finite Element	PDF	$8.362 \times 10^{-5}$ ( $7.014 \times 10^{-3}$ )	$-3.335 \times 10^{-5}$ ( $8.618 \times 10^{-4}$ )	$-1.878 \times 10^{-5}$ ( $2.663 \times 10^{-5}$ )	5.62
	PDF (interp.)	$1.592 \times 10^{-4}$ ( $6.353 \times 10^{-3}$ )	$-2.426 \times 10^{-5}$ ( $8.804 \times 10^{-4}$ )	$-1.809 \times 10^{-5}$ ( $3.080 \times 10^{-5}$ )	1.26
	CDF	$1.405 \times 10^{-4}$ ( $6.682 \times 10^{-3}$ )	$-2.345 \times 10^{-5}$ ( $8.625 \times 10^{-4}$ )	$-1.645 \times 10^{-5}$ ( $2.773 \times 10^{-5}$ )	7.15
	CDF (interp.)	$1.743 \times 10^{-4}$ ( $6.752 \times 10^{-3}$ )	$-2.107 \times 10^{-5}$ ( $8.495 \times 10^{-4}$ )	$-1.614 \times 10^{-5}$ ( $3.173 \times 10^{-5}$ )	1.38

Table 8: Mean difference between EML estimates of the parameters of the OU model and the estimates obtained from a variety of numerical procedures. The average is taken over 2000 different samples of the process  $dX = \alpha(\beta - X)dt + \sigma dW$  with parameters  $\alpha = 0.2$ ,  $\beta = 0.08$  and  $\sigma = 0.03$ . The standard deviation of the difference between the estimates is shown in parentheses. The average EML parameter estimates over the 2000 samples of  $N = 500$  transitions are  $\alpha = 0.3101$ ,  $\beta = 0.0794$  and  $\sigma = 0.0301$ . The average EML parameter estimates over the 2000 samples of  $N = 1000$  transitions are  $\alpha = 0.2467$ ,  $\beta = 0.0798$  and  $\sigma = 0.0300$ .

## 8 Empirical application

There is a large literature documenting the link between the instantaneous short-term interest rate and the pricing of zero-coupon bonds for various terms to maturity. The classic one-factor model of the term structure is based on the assumption that the instantaneous interest rate  $r(t)$  evolves in accordance with

$$dr = \alpha(\beta - r) dt + \sigma r^\gamma dW \quad (55)$$

where  $\alpha$  (speed of adjustment),  $\beta$  (the mean interest rate),  $\sigma$  (volatility control) and  $\gamma$  (the so-called levels effect) are the parameters to be estimated. Notwithstanding any theoretical shortcomings in this general approach, estimating the parameters of (55) is a worthwhile and challenging task with which to illustrate the methods outlined here and reach some tentative conclusions on the magnitudes of the parameters of interest. In this section the CDF-based finite-element procedure, with and without interpolation, will be used to estimate the parameters of the generalised CIR<sup>7</sup> model and also the special case of  $\gamma = 0.5$ , which is the CIR model discussed previously. The data are monthly average observations of the U.S. 3-month Treasury Bill rate from August 1982 to November 1998 obtained from the Federal Reserve Economic Database<sup>8</sup> and are plotted in Figure 2. This data set has previously been used by Hurn *et al.* (2003) in the estimation of the parameters of SDEs.

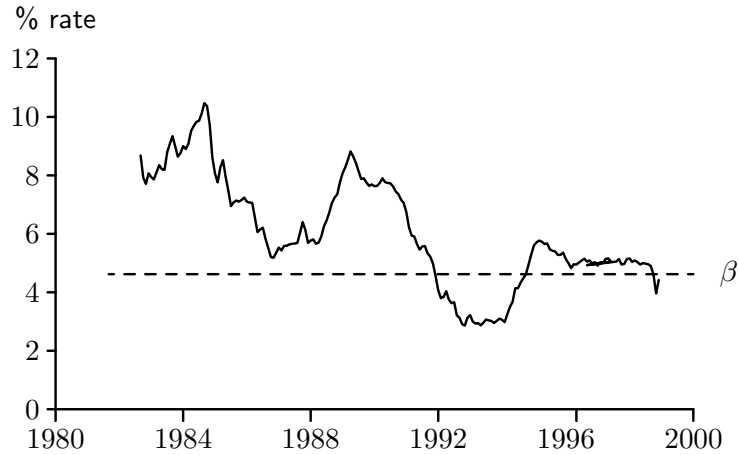


Figure 2: Monthly average of the U.S. 3-month Treasury Bill rate from August 1982 to November 1998.

As the true parameters are unknown in this case, no quantitative conclusions can be drawn about the accuracy of the parameter estimates. On the other hand, as the closed-form expression for the transitional PDF is known for the special case of the CIR process, EML may again be used to estimate the parameters in order to provide a benchmark for comparison. The results, presented

<sup>7</sup>This model is also known as the constant elasticity of variance model

<sup>8</sup>The data may be downloaded from the URL <http://www.stls.frb.org/fred/>

in Table 9, clearly demonstrate the computational benefits of the interpolation idea despite the fact that the number of observations (196) in the data set is small. The interpolation idea reduces the computational time of the CDF-based finite-element procedure by roughly half without any degradation in the accuracy of the parameter estimates.

	$\alpha$	$\beta$	$\sigma$	$\gamma$	Time (sec)
Generalised CIR Process					
Finite element	0.1554	0.0438	0.0514	0.6443	121.31
Finite element (interp.) ( $M = 90$ )	0.1554	0.0438	0.0514	0.6443	59.13
CIR Process					
EML	0.1559	0.0438	0.0341	—	0.02
Finite element	0.1557	0.0438	0.0341	—	64.19
Finite element (interp.)( $M = 90$ )	0.1557	0.0438	0.0341	—	30.37

Table 9: Parameter estimates for CIR and Generalised CIR SDEs using Treasury Bill rate data. In the FE scheme the discretisation intervals are  $n = 1000$  and  $\Delta_t = 1/1200$  for state space and time respectively.

This application highlights the need to choose carefully the number of interpolating nodes. The choice of  $M = 90$  for both models was made with reference to a quick calibration exercise. First, the parameters of the SDEs were estimated using discrete maximum likelihood (DML). Second, the log-likelihood values provided by the interpolated finite difference scheme at the DML parameters were examined as  $M$  was increased from 50 in steps of 10. The chosen values reflect the points at which the log-likelihoods were insensitive (to four decimal places) to the addition of more nodes.

## 9 Conclusion

This paper has explored three different ways of improving the numerical solution of the Fokker-Planck equation and the associated task of estimating the parameters of SDEs by maximum likelihood. It is demonstrated that a reformulation of the Fokker-Planck equation in terms of the transitional CDF circumvents the difficulties presented by the delta function initial condition for the traditional PDF specification of the problem and improves the robustness of the numerical procedure substantially. Furthermore, it is shown that the finite-element approach to solution of the partial differential equation delivers more accurate parameter estimates with less computational effort than equivalent procedures using the method of finite differences. Finally, an interpolation procedure is developed that relies on the fact that many of the SDEs in finance are time homogeneous. In this situation information about the value of the transitional PDF for a grid of initial states can be used to solve, by interpolation, for the value of the transitional PDF for any initial state.



The robust conclusion of this paper is that an estimation procedure that incorporates all of these contributions, that is, one that solves a CDF-based specification of the Fokker-Planck equation numerically using the method of finite-elements and also takes advantage of the interpolation technique, has much to recommend it. It is accurate, computationally efficient, robust and completely generic in the sense that once coded new problems can be handled with trivial modifications of the code.

## References

- Aït-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*, **9**, 385-426.
- Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, **70**, 223-262.
- Bakshi, G. and Ju, N. (2005). A Refinement to Aït-Sahalia's (2002) "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach". *Journal of Business*, **78**, 2037-2052.
- Bibby, B.M. and Sørensen, M. (1995). Martingale Estimation Functions for Discretely Observed Diffusion Processes. *Bernoulli*, **1**, 17-39.
- Brandt, M.W. and Santa-Clara, P. (2002). Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. *Journal of Financial Economics*, **63**, 161-210.
- Chacko, G. and Viceira, L.M. (2003). Spectral GMM Estimation of Continuous-Time Processes. *Journal of Econometrics*, **116**, 259-292.
- Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, **47**, 1209-1227.
- Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, **53**, 385-407.
- Durham, G.B. and Gallant, A.R. (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business and Economic Statistics*, **20**, 297-316.
- Elerian, O. (1998). A Note on the Existence of a Closed Form Conditional Transition Density for the Milstein Scheme. *Working Paper, Nuffield College, Oxford University*.
- Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood Inference for Discretely Observed Non-linear Diffusions. *Econometrica*, **69**, 959-993.
- Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics*, **19**, 177-191.
- Feller, W. (1951). Two Singular Diffusion Problems. *Annals of Mathematics*, **54**, 173-182.
- Florens-Zmirou, D. (1989). Approximate Discrete-Time Schemes for Statistics of Diffusion Processes. *Statistics*, **20**, 547-557.

- Gallant, A.R. and Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, **12**, 657-681.
- Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, **8**, 85-118.
- Hansen, L.P. and Scheinkman, J.A. (1995). Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica*, **63**, 767-804.
- Hurn, A.S. and Lindsay, K.A. (1999). Estimating the Parameters of Stochastic Differential Equations. *Mathematics and Computers in Simulation*, **48**, 373-384.
- Hurn, A.S., Lindsay, K.A. and Martin, V.L. (2003). On the Efficacy of Simulated Maximum Likelihood for Estimating the Parameters of Stochastic Differential Equations. *Journal of Time Series Analysis*, **24**, 45-63.
- Hurn, A.S., Jeisman, J. and Lindsay, K.A. (2006). Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations. *National Centre for Econometric Research Discussion Paper*.
- Jensen, B. and Poulsen, R. (2002). Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques. *Journal of Derivatives*, **9**, 18-32.
- Jiang, G.J. and Knight, J.L. (2002). Estimation of Continuous-Time Processes via the Empirical Characteristic Function. *Journal of Business and Economic Statistics*, **20**, 198-212.
- Jones, C.S. (1998). A Simple Bayesian Method for the Analysis of Diffusion Processes. *Working Paper, Simon School of Business, University of Rochester*.
- Kessler, M. and Sørensen, M. (1999). Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process. *Bernoulli*, **5**, 299-314.
- Pedersen, A.R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, **22**, 55-71.
- Roberts, G.O. and Stramer, O. (2001). On Inference for Partially Observed Non-linear Diffusion Models using the Metropolis-Hastings Algorithm. *Biometrika*, **88**, 603-621.
- Shoji, I. and Ozaki, T. (1998). Estimation for Nonlinear Stochastic Differential Equations by a Local Linearization Method. *Stochastic Analysis and Applications*, **16**, 733-752.
- Singleton, K.J. (2001). Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function. *Journal of Econometrics*, **102**, 111-141.
- Sørensen, M. (2000). Prediction Based Estimating Functions. *Econometrics Journal*, **3**, 123-147.

Sundaresan, S.M. (2000). Continuous-time Methods in Finance: A Review and an Assessment. *Journal of Finance*, **55**, 1569-1622.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, **5**, 177-188.

## A Integrals involving basis functions

This appendix develops the integral results used in the derivation of equations (29), (30), (31), (40), (41) and (42). The important point to note is that the value of these integrals is the sum of contributions from individual elements. Consider the contribution made by the single element occupying  $[x_{j-1}, x_j]$  in the case of the family of triangular basis functions

$$\psi_j(x) = \begin{cases} \frac{x - x_{j-1}}{x_j - x_{j-1}} & x \in [x_{j-1}, x_j] \\ \frac{x_{j+1} - x}{x_{j+1} - x_j} & x \in [x_j, x_{j+1}] \\ 0 & \text{otherwise.} \end{cases} \quad (56)$$

Only  $\psi_{j-1}(x)$  and  $\psi_j(x)$  take a non-zero value for this element. Let  $p, q, r$  and  $s$  be non-negative integers, then the coefficient of any contribution to the finite-element algorithm arising from this element is a special case of the generic form

$$I(p, q, r, s) = \int_{x_{j-1}}^{x_j} \psi_{j-1}^p(x) \psi_j^q(x) \left( \frac{d\psi_{j-1}}{dx} \right)^r \left( \frac{d\psi_j}{dx} \right)^s dx. \quad (57)$$

Under the change of variable  $x = x_{j-1} + \lambda(x_j - x_{j-1})$ , the basis functions  $\psi_{j-1}(x)$  and  $\psi_j(x)$  become respectively  $(1 - \lambda)$  and  $\lambda$  and the expression for  $I(p, q, r, s)$  simplifies to give

$$I(p, q, r, s) = \frac{(-1)^r}{(x_j - x_{j-1})^{r+s-1}} \int_0^1 (1 - \lambda)^p \lambda^q d\lambda = \frac{(-1)^r B(p+1, q+1)}{(x_j - x_{j-1})^{r+s-1}} \quad (58)$$

where  $B(m, n)$  denotes the usual Beta function. The construction of the finite-element solution of the Fokker-Planck equation requires the evaluation of integrals in which the integrand is a product involving either two or three basis functions.

### Integrals involving products of two basis functions

The use of the finite-element procedure to solve the Fokker-Planck equation requires the evaluation of two such integrals. The general result (58) may first be used to show that

$$\begin{aligned} \int_{x_{j-1}}^{x_j} \psi_{j-1}(x) \psi_j(x) dx &= (x_j - x_{j-1}) B(2, 2) = \frac{x_j - x_{j-1}}{6}, \\ \int_{x_{j-1}}^{x_j} \psi_j^2(x) dx &= (x_j - x_{j-1}) B(3, 1) = \frac{x_j - x_{j-1}}{3}. \end{aligned} \quad (59)$$

In the case in which the finite-element method is used to solve the partial differential equation describing the evolution of the transitional CDF, the construction of the transitional PDF from the transitional CDF requires the evaluation of

$$\begin{aligned}\int_{x_{j-1}}^{x_j} \frac{d\psi_j(x)}{dx} \psi_j(x) dx &= - \int_{x_{j-1}}^{x_j} \frac{d\psi_{j-1}(x)}{dx} \psi_j(x) dx = B(1, 2) = \frac{1}{2}, \\ \int_{x_{j-1}}^{x_j} \frac{d\psi_j(x)}{dx} \psi_{j-1}(x) dx &= - \int_{x_{j-1}}^{x_j} \frac{d\psi_{j-1}(x)}{dx} \psi_{j-1}(x) dx = B(1, 2) = \frac{1}{2}.\end{aligned}\tag{60}$$

### Integrals involving products of three basis functions

The use of the finite-element procedure to solve the Fokker-Planck equation requires the evaluation of two such integrals. The integrand of the first class of integral is the product of two basis function and the derivative of a basis function. In this case, the general result (58) may be used to show that

$$\begin{aligned}\int_{x_{j-1}}^{x_j} \psi_{j-1}^2 \frac{d\psi_j}{dx} dx &= - \int_{x_{j-1}}^{x_j} \psi_{j-1}^2 \frac{d\psi_{j-1}}{dx} dx = B(3, 1) = \frac{1}{3}, \\ \int_{x_{j-1}}^{x_j} \psi_{j-1} \psi_j \frac{d\psi_j}{dx} dx &= - \int_{x_{j-1}}^{x_j} \psi_{j-1} \psi_j \frac{d\psi_{j-1}}{dx} dx = B(2, 2) = \frac{1}{6}, \\ \int_{x_{j-1}}^{x_j} \psi_j^2 \frac{d\psi_j}{dx} dx &= - \int_{x_{j-1}}^{x_j} \psi_j^2 \frac{d\psi_{j-1}}{dx} dx = B(1, 3) = \frac{1}{3}.\end{aligned}\tag{61}$$

The integrand of the second class of integral is the product of one basis function and two derivatives of a basis function. The general result (58) may again be used to show that

$$\begin{aligned}\int_{x_{j-1}}^{x_j} \psi_j \left( \frac{d\psi_j}{dx} \right)^2 dx &= - \int_{x_{j-1}}^{x_j} \psi_j \frac{d\psi_{j-1}}{dx} \frac{d\psi_j}{dx} dx = \frac{B(1, 2)}{(x_j - x_{j-1})} = \frac{1}{2(x_j - x_{j-1})}, \\ \int_{x_{j-1}}^{x_j} \psi_{j-1} \left( \frac{d\psi_j}{dx} \right)^2 dx &= - \int_{x_{j-1}}^{x_j} \psi_{j-1} \frac{d\psi_j}{dx} \frac{d\psi_{j-1}}{dx} dx = \frac{B(2, 1)}{(x_j - x_{j-1})} = \frac{1}{2(x_j - x_{j-1})}.\end{aligned}\tag{62}$$